Topics in nonparametric identification and estimation

Hubner, Stefan

Publication date:
2016

Document Version
Publisher's PDF, also known as Version of record

Topics in Nonparametric Identification and Estimation

Proefschrift

ter verkrijging van de graad van doctor aan Tilburg University op gezag van de rector magnificus, prof. dr. E.H.L. Aarts, in het openbaar te verdedigen ten overstaan van een door het college voor promoties aangewezen commissie in aula van de Universiteit op *vrijdag 18 november 2016 om 14.00 uur* door

Stefan Hubner

geboren op 14 juni 1985 te Bruck an der Mur, Oostenrijk

Promotor

prof. dr. Arthur van Soest

Copromotor

dr. Pavel Čížek

Promotiecommissie

prof. dr. Jaap Abbring

prof. dr. Laurens Cherchye

prof. dr. Frederic Vermeulen

Acknowledgements

The six years that led to this thesis have been an incredible journey for me, both intellectually and personally. I do not want to pretend that the result was a solo effort and far be it from me to take all the credit for it. I am certain that without the help of a range of extraordinary individuals who I owe a debt of gratitude, this would not have been possible and I would not be where I am today.

First and foremost my deepest appreciation goes to Arthur. I approached you after the Research Master to express my interest in a project you were coordinating and you gave me the chance to be a part of it. As my advisor you not only devoted an incredible amount of time to me, but also showed tremendous patience when I was trying to order my thoughts on your whiteboard during our meetings. I very much appreciate the freedom you gave me to work on my own ideas. It can be difficult to distribute the workload of a doctoral thesis over the period of several years. Our weekly meetings and the occasional nudge, when I lost focus or got carried away with other ideas, helped me to stay on track and finally finish this thesis. Further, I want to express my gratitude to Pavel, who was actually my first advisor from a chronological perspective. You always had an open door for me and our meetings would sometimes last up to half a day. I have learned a lot from you.

I am deeply grateful to Stefan for hosting me during a research visit at Boston College at a very important stage of my PhD. With the job market around the corner not only did you support my applications, but also managed to boost the confidence I had in myself and my research. You welcomed me into your home, introduced me to a lot of interesting people and we had many good conversations, which were not only research-related. During my visit I also met Arthur Lewbel with whom I had fruitful discussions over a cup of gunpowder tea, that he had brought from his latest trip to the Far East. I am very grateful for your constructive input and support.

In addition to this I want to thank my doctoral committee Frederic, Jaap and Laurens, for the time and effort you put into this task and for your comprehensive and helpful comments. It is an honor for me to draw from the expertise of such a distinguished group of researchers, who not only contributed to, but also fundamentally shaped the literature. Moreover, I very much appreciate the assistance I received from the job market committee, in particular Otilia for preparing and actively promoting me in this critical process.

This thesis closes an important chapter of my academic career. I am grateful that I encountered so many interesting people from very heterogeneous backgrounds, with whom I had not only great discussions on a variety of topics, but also a lot of fun and just a great time. At this place I want to mention my officemate Chen from whom I have learned a lot about the Chinese culture, as well as the ever-evolving lunch group including Ahmadreza, Alaa, Bas, Frans, Gyula, Jarda, Maria, Marieke, Marleen, Mitzi, Nick, Olga, Sara and Rasa. I particularly want to thank Jan for occasionally pushing me out of my comfort zone and for sharing all of his most recent experiences with me. I will also cherish the often heated arguments I had with Sybren about a variety of philosophical topics and the many discussions I had with Bas, which initially had been research-related, but were gradually replaced by our common hobby of cycling and in particular the aerodynamics thereof. I also want to thank the rest of the cycling group Liz, Mario and Sebastian for beautiful, yet challenging rides together, as well as the windsurfing crew in Hoek van Holland for many memorable moments on the water.

My heartfelt appreciation goes to my family. My parents Leopold and Marion for always encouraging and supporting me, my sister Sabrina for reminding me to take a step back and reflect every once in a while and not always be too hard on myself, Axel for keeping me grounded to the world outside of academia and last but not least my grandpa Karl, who awakened my interest in science and technology at a young age.

Finally, no words can express my gratitude to Astrid for your love and support. I cannot imagine how I could have done this without the strength and stability you gave me for the last almost ten years. Thank you for putting up with my sometimes disorganized ways and my unwillingness to commit to plans by being the J to my P.

Tilburg, August 2016

Contents

Introduction **xi**

1 Collective Households with Heterogeneous Agents **1**

1.1 Introduction 1

1.2 Specification and Economic Restrictions 5

1.3 Identification 10

1.4 Estimation of Demands and the Sharing Rule 18

1.5 Monte Carlo Simulations 22

1.6 Empirical Application 30

1.7 Conclusion 38

1.A Auxiliary Results and Proofs 39

1.B Further Application Results 53

1.C Time-use and Consumption Categories 55

2 The Collective Axiom in a Heterogeneous Population **57**

2.1 Introduction 57

2.2 Specification 61

2.3 Encoding 66

2.4 Computation and Inference 70

2.5 Simulations 74

2.6 Conclusion 77

2.A Algorithms 78

3 Smooth Transition GARCH Models **79**

3.1 Introduction 79

3.2 Model Specification 82

3.3 Estimation Procedure 85

3.4 Asymptotic Results 89

3.5 Monte-Carlo Simulations 92

3.6 Empirical Application 101

3.7 Conclusion and Outlook 106

3.A Auxiliary Results and Proofs 107

Bibliography **123**

List of Figures

1.1 Data-Generating Process: Public 25

1.2 Surface plot: Elasticities ranging from -1.5 to $+1.5$ for $\tau \in \{0.1, 0.5, 0.9\}$ . . 35

2.1 Three intersecting budget sets $B_{red}, B_{blue}, B_{green}$ with three goods 67

2.2 Power function for $N = 1{,}500$ (l.h.s) and $N = 3{,}000$ (r.h.s.) 76

2.3 Type I error for $n_0 = 5$ (l.h.s) and $n_0 = 2$ (r.h.s.) worst-case paths 77

3.1 Logistic, Linear and Threshold Transition Function on $[F_{\xi_t}^{-1}(0.05), F_{\xi_t}^{-1}(0.95)]$ 93

3.2 DAX returns in % and its corresponding 5% and 95% ANST-GACQ estimates 103

3.3 DAX: Convergence for Coverage Tests, 5% (l.h.s) and 25% (r.h.s.) 106

List of Tables

1.1 Private/Public Almost Ideal Demand Systems, $\tau = 0.5$, with integration . 27

1.2 Private/Public Almost Ideal Demand Systems, $\tau = 0.5$, without integration 28

1.3 Private/Public Almost Ideal Demand Systems, $\tau = 0.9$, with integration . 29

1.4 Private/Public Almost Ideal Demand Systems, $\tau = 0.9$, without integration 30

1.5 Wife, Husband and Household Descriptive Statistics (Raw) 32

1.6 Wife, Husband and Household Descriptive Statistics (Model) 34

3.1 Sample Size 95

3.2 Order of Approximation: $m = \lceil cn^{1/4} \rceil$ 96

3.3 $\delta \in (0,1)$ where central $0.5 \pm \frac{\delta}{2}$ quantiles are not estimated 96

3.4 Model comparison: GACQ and GARCH with Normal errors 97

3.5 Model comparison: GACQ and GARCH with Student's t(4) errors 98

3.6 Model Comparison for re-centered, mirrored Gumbel errors 99

3.7 Outliers 100

3.8 Misspecified transition functions: (G, G_0) for GACQ and GARCH-N . . . 101

3.9 Coefficients and standard errors for 5%-VaR ANST-GACQ estimates . . . 103

3.10 Coverage and Test statistics: Comparison to GARCH 105

3.11 Coverage and Test statistics: Comparison to single regime 105

Introduction

Moore's law states that the number of transistors on a given area of a microprocessor will approximately double every two years. While Gorden Moore, one of the founders of Intel, who made this rather ambitious prediction in the April 1965 edition of Electronics magazine was referring to the foreseeable future, it is remarkable how accurately the development of technology follows this law even up to this day. To put the magnitude of this into perspective, at the time it was possible to put less than one thousand transistors on a silicon chip, whereas today the latest CPU and even GPU models host tens of billions thereof, even without taking into account the possibility of using many of them in parallel. In addition to this, not only did computers become faster but also a lot cheaper. Clearly this development has revolutionized many fields of science, with Econometrics being no exception.

In the earlier days, Applied Economists and Econometricians mostly focused on identifying parameters of an economic model defined as a single equation or a system of multiple equations in combination with a parametric distribution function to model measurement error or unobserved heterogeneity [Koopmans & Reiersol, 1950; Hurwicz, 1950; Fisher, 1966; Rothenberg, 1971]. Being less and less restricted by computational capacities due to the aforementioned advancement of computer hardware, the focus gradually shifted from parametric to nonparametric approaches, which are concerned with the identification and estimation of functions and distributions not belonging to a parametric family. Nonparametric models have the advantage of requiring less assumptions on the structure of an economic model. For example a researcher might be sympathetic to a rationality assumption imposed on individuals, but at the same time does not believe that the shape of Engle curves can be globally represented by a pre-specified parametric function for all prices and income levels, but rather believes in a more general relationship between demands and endowment. A different example which is often studied using a nonparametric approach and of fundamental interest in this thesis is the modelling of unobserved heterogeneity. Coming back to the previous example, it is widely accepted that individuals are rational utility optimizers, however, even after controlling for different incomes and a variety of demographic statistics, researches find a substantial amount of heterogeneity of demands which is left unexplained [Blundell & Stoker, 2007]. Due to the lack of results on identification of functions that are non-separable with respect to unobserved preference heterogeneity parameters, without substantially restricting utility functions and thus preferences [Lewbel, 2001] it was difficult to allow for a deeper interpretation of this heterogeneity in demands other than measurement errors.

Roehrig [1988] was the first to study non-parametric identification of simultaneous equation models with non-additive error terms based on techniques developed earlier by Brown [1983]. However, in a comment Benkard & Berry [2006] provide a counterexample showing that the conditions for identification of structural functions they provide for such a setting are not sufficient. Later on, by exploiting information about quantiles of the distribution of the non-separable unobserved variable in a single equation context Matzkin [2003]; Altonji & Matzkin [2005] show how to non-parametrically identify the structural demand function using a monotonicity assumption. Chesher [2003] extends this to the context of a system of non-linear and non-additive equations using a triangularity restriction. While identification and estimation of moments of the structural functions were studied previously in a semiparametric context Powell [1994]; Härdle et al. [1991]; Härdle & Stoker [1989] and later in a nonparametric one by Matzkin [2008]; Imbens & Newey [2009], these approaches made it possible to give the distribution of preference parameters a meaningful interpretation by being able to interpret the latter as a range of types which is then mapped to the quantile of the structural (demand) function. As a consequence, under certain integrability restrictions, this allows welfare analysis of a heterogeneous population since it is possible to link the structural demand function to a cardinalization of utility functions, both indexed by a vector of preference parameters, the distribution of which is observed.

The first chapter of this thesis builds upon these techniques when considering non-parametric identification and estimation of structural components of the collective household consumption model [Chiappori, 1988, 1992] in which household members bargain over their consumption choices and are assumed to reach a Pareto efficient outcome. Particular attention is given to non-separable unobserved heterogeneity in the reduced form demands that arises from the underlying aggregate decision process. For this, I derive necessary and sufficient conditions the model's primitives have to satisfy in order to ensure nonparametric point-identification of the conditional sharing rule, a central component of the collective model that determines the allocation of endowment among household members. A crucial condition is the existence of information on the intra-household allocation of consumption in the considered dataset. The Dutch LISS Panel (CentERData) is one example of such a dataset for which this approach is feasible. In addition to showing nonparametric identification of structural components of this model, I also develop a nonparametric conditional quantile estimation procedure [Koenker, 2000, 2005] based on smoothing techniques [Chaudhuri, 1991; Yu & Jones, 1998] and derive its asymptotic properties. In a Monte-Carlo experiment I study both the finite sample behaviour of the proposed estimation procedure and compare it to a naive non-parametric estimator. I also specify a collective labour supply model using the LISS panel and estimate demands and conditional sharing rules for different parts of the distribution of preference characteristics and find that there is indeed a significant amount of heterogeneity even up to the extent that the sign for some elasticities change with respect to the considered quantile of the taste distribution.

It is not always necessary to recover structural functions. To test implications of a model and falsify them in the sense of Popper [1934], it is often sufficient to identify only certain characteristics of the distribution of structural components of an economic model or functionals thereof. This characteristics could be moments as in Härdle et al. [1991]; Blundell & Powell [2003]; Imbens & Newey [2009] who consider conditional means of structural functions and derivatives thereof, Hoderlein [2011] who additionally considers conditional variances and, finally, Hoderlein & Mammen [2007]; Dette et al. [2016] who exploit information about local averages of conditional quantiles of structural functions. Another interesting case is the one where relevant characteristics of the distribution of structural components are only set identified rather than point-identified (see Chesher & Rosen [2014] for a definition), which leads to the vastly growing literature of partial identification (see for example Tamer [2010] for a comprehensive survey of the relevant methods). One example is the case in which, for example due to the lack of more informative datasets, one has to reason via marginal distributions because the joint distribution is not observed. Fréchet [1951] first studied this problem by bounding joint probabilities using copulas which was later formalized in a more general Econometric setting by Manski [1995, 2003, 2007].

Chapter two considers the so-called collective axiom of revealed preference in the context of a partial identification setting. Once more I study the collective household consumption model which assumes individual rationality and Pareto-efficient bargaining within the household. This imposes restrictions on individual and aggregate demands under different budget situations, which are characterized by prices and endowments. This set of conditions is known as the Collective Axiom of Revealed Preference [Cherchye et al., 2007, 2009]. In this chapter I show how one can exploit data from single households in such a revealed preference context to further open the black box of intra-household decision making. This approach requires the structural assumption that preferences are stable with respect to different household compositions. In particular, I propose a non-parametric test of this assumption which allows for unobserved heterogeneity both with respect to preferences and intra-household bargaining. Making use of a finite-dimensional characterization of hypothetical household types, using the idea of stochastic revealed preferences [McFadden & Richter, 1991; McFadden, 2005], I show how to construct a test-statistic based on observing only marginal distributions of consumption choices for single and couple households, respectively, by partially identifying the joint distribution. Finally a simulation study is conducted providing evidence that the test has power against the alternative hypothesis of non-stable preferences and shows that it is correctly sized under different worst-case scenarios.

The third chapter of this thesis is co-authored by Pavel Čížek. It is best classified as a semiparametric identification approach [Powell, 1994] with elements of nonparametric estimation. Lewbel [2016] defines a semi-parametric model in a way that the estimation of the parameters requires coping with the nonparametric components of the model. While methodologically this chapter is positioned well within the broad topic of nonparametric identification, from a content-related point of view it deviates from the first two to the extent that it considers a time series framework, rather than microeconomic data. In particular, we introduce non-linearities to the Conditional Autoregressive Value at Risk (CAViaR) model [Engle & Manganelli, 2004]. This is motivated by the fact that financial time series often exhibit asymmetric dynamic behaviour with respect to past shocks not only in their conditional means but also in their conditional volatilities. The focus is on the latter, in particular if the main interest is not in the conditional volatility itself but rather in the conditional quantile. We propose a general autoregressive conditional quantile model that allows for asymmetric behaviour and is robust to distributional assumptions. Our model can be viewed as an extension of the CAViaR model in which we allow for two parameter regimes. We consider both a threshold and a smooth transition version by using a parametrized transition function with a location and a scale parameter. We show that this transition function is nonparametrically identified. Further we propose an estimation procedure employing a sieve estimator [Chen, 2008], which belongs to the class of nonparametric estimators, in order to estimate conditional volatilities in a first stage using composite quantile regression. These estimates are then used in the generalized autoregressive conditional quantile estimation in the second stage. We show that our estimator is consistent and asymptotically normal. Our simulations indicate that prediction errors can be improved over both the smooth transition General Autoregressive Conditional Heteroskedasticity model (GARCH) and the single regime CAViaR model, which are considered to be the standard methods. In an empirical application we investigate asset returns of USD/GBP and the German equity index (DAX). We find strong evidence that exchange rate between the US Dollar and the British Pound is governed by two parameter regimes.

CHAPTER 1

Collective Households with Heterogeneous Agents

Introduction

Since the seminal contribution of Chiappori [1988, 1992] the collective household model has found increasing popularity in consumption and labour supply settings. It is argued that one should prefer a model in which each individual acts as a decision maker, as opposed to the traditional unitary model which assumes that the households acts as if it would follow one set of goals. Indeed there are many realms of applications, e.g. female labour supply, welfare economics and family economics, for which the collective model provides additional insight into the household decision making process. While the micro-economic foundation of the collective model for a representative consumer is well developed, very little is known about how one can incorporate the notion of unobserved heterogeneity into the model. It is generally accepted that the assumption of homogeneity does not hold in many domains of applications, since group statistics often still exhibit a significant amount of variation [Blundell & Stoker, 2007]. In the context of the collective model homogeneity would require individual preferences and the within-household allocation process, to be identical given observed characteristics, such as prices, endowment and demographic variables. A rather ad-hoc way this issue is usually dealt with, is to solve the model for a representative consumer, i.e. as if there was no unobserved heterogeneity, and then include an additive and heteroskedastic but separable error term in the system of demand equations. From a structural point of view, such an additive error term can be interpreted as a measurement error. In other words, representative consumers maximize a common utility function subject to their budget constraint but observed demands are subject to recording errors or optimization errors. On the other hand, if the error terms are supposed to capture unobserved preference heterogeneity, an additive structure poses very strong restrictions on the individuals' preferences and the aggregation process. Even within the more traceable unitary setting, in general a stochastic demand system that admits a broad class of underlying utility functions, turns out to be non-separable in the error terms [Brown & Walker, 1989; Lewbel, 2001]. No such result seems to exist for the collective model. To the knowledge of the author, the following contributions exist that allow for structural heterogeneity within the collective model. Chiappori et al. [2012] considers the empirical content of Nash bargaining, which in its essence is a collective model however differs from the usual collective model in the sense that it incorporates all common bargaining axioms instead of only Pareto efficiency. Chiappori & Kim [2013] use a collective household setting and show identification of derivatives of the sharing rule with respect to a distribution factor. This paper considers representative customers when it comes to taste preferences (up to a measurement error), but allows for a structural additive error in the sharing rule which can be interpreted as unobserved bargaining heterogeneity. Matzkin & Perez-Estrada [2011] on the other hand consider additive sub-utility functions capturing household taste preferences using a scalar household-wide taste shock. Dunbar et al. [2013b] consider completely random resource shares. They show that under certain preference restrictions and the existence of a number of distribution factors, the joint distribution of resource share levels is identified.

In this paper we attempt to take another step towards getting a better understanding of how one can non-parametrically identify and estimate structural components of the model, if unobserved heterogeneity is introduced not only as an error term in the demands but in the models primitives. We will allow for taste shocks which capture individuals' preferences for each good as well as unobserved heterogeneity in the members' bargaining power, which will have an effect on the within-household allocation process. We do not require an additively or multiplicatively separable error term, but will rather provide necessary and sufficient conditions on the models primitives that will ensure invertibility of demands with respect to the error terms. It is important to note that our approach is fundamentally nonparametric, which allows us to make conclusions about a broad class of utility functions and aggregation rules rather than a specific parametrization, ruling out the possibility of misspecification bias. In addition to this, in order to keep our results as general as possible, we allow for an arbitrary number of group members, consuming both private and public goods. One of the main difficulties within the context of the collective model is to identify levels of the conditional sharing rule, which captures how household endowment is distributed among the members. While there exists a whole literature concerned about strategies to recover the latter from aggregate consumption information, see for example Browning et al. [2013]; Dunbar et al. [2013a] who impose restrictions on preferences, or Cherchye et al. [2015] who propose a set identification strategy using the collective axiom of revealed preferences, it is not the purpose of this paper to provide a new strategy that recovers the sharing rule, but rather to show its identification in the presence of unobserved heterogeneity, once it is recovered. For this reason this paper assumes fully assignable consumption information of household members. There exists a range of datasets, that are feasible for such an identification strategy including the time-use and consumption module of the *Dutch LISS panel* [Cherchye et al., 2012], the *Danish Household Expenditure Survey* [Bonke & Browning, 2015], a survey of the *Italian International Center of Family studies* [Menon et al., 2012] and for time-use data the *UK Time Use Survey 2000* [Browning & Gortz, 2012]. This allows us to separate the endowment allocation problem resulting in the conditional sharing rule and the individual consumption decision problems, not only when considering the underlying economic structure but also within our empirical setting in which we will estimate conditional quantiles of the conditional sharing rule. While conditional quantiles of this quantity are identified immediately, the main challenge and the main focus of this paper is to relate them to the distribution of taste and bargaining shocks representing unobserved heterogeneity which will ultimately allow us to estimate the (causal) effect of a (policy relevant) impulse on the distribution of the conditional sharing rule, as well as public and private good consumption. One could think of policies concerning tax transfers as for example family splitting, in which spouses pool their income for taxation giving them tax benefits in a progressive tax system or the question who of the spouses receives child benefit payments. Since one might expect that responses to such policies might differ across the population, looking at conditional quantiles can provide very valuable insights to policy makers.

As opposed to the aforementioned contributions of Chiappori et al. [2012]; Chiappori & Kim [2013]; Dunbar et al. [2013b] our approach will largely draw from the existing non-parametric identification literature. The restrictions we will derive on the models' structure, will be a consequence of higher level restrictions imposed on the demand functions that are required for these general non-parametric identification results to apply. Functions, both scalar and vector valued, that are non-separable with respect to error terms have been heavily studied in the econometrics literature. See for example Brown [1983]; Roehrig [1988] who consider identification of conditional means, Chesher [2003]; Matzkin [2003]; Altonji & Matzkin [2005]; Imbens & Newey [2009] who focus on triangular and invertible models considering conditional quantiles or Hoderlein & Mammen [2007, 2009] who provide identification results based on local average structural derivatives. These results have been applied to the unitary consumption model, but not yet to the collective one. Perhaps the closest paper to this one for the unitary model is Beckert & Blundell [2008] who provide necessary and sufficient conditions which marginal rates of substitutions have to satisfy to ensure invertibility of Walrasian demands. Blundell et al. [2013] and Blundell et al. [2014] propose estimation procedures for non-separable demand functions under Slutsky and Revealed Preference restrictions, respectively. The main advantage of the latter approaches is that rationality restrictions are imposed while at the same time the flexibility of non-parametric estimation is maintained. All the above approaches have in common that they can be sub-summed under the monotonicity or invertibility paradigm. While this is a more comprehensive approach that allows for full point-identification of demands, it requires the researcher to impose more structure on the underlying model. On the other end of the scale there is identification of (local) averages. These approaches permit very general forms of unobserved heterogeneity and are mostly used for testing implications of individual rationality. Härdle et al. [1991] for examples investigates the empirical content of the law of demand imposing the metonymy hypothesis, whereas Hoderlein [2011]; Haag et al. [2009]; Dette et al. [2016] test the empirical content of Slutsky symmetry and negative-definiteness and Hoderlein & Stoye [2014] tests the weak axiom of revealed preferences in a unitary setting.

The method we follow in this paper will be closer to the invertibility strand of the literature, although there will be some elements of average derivative estimation. While we will be able to achieve full point-identification of demand functions with all its advantages, it is obvious that some assumptions have to be met by the demand system in order for them to apply. The challenge is to connect these assumptions to the models' primitives to study the restrictions they impose on the underlying structure.

The outline of the paper is as follows. In Section 1.2 we will briefly discuss the economic structure of the collective model and derive its key restrictions that we will use to link demands to the underlying utilities and aggregation process. Based on these results, in Section 1.3 we will derive restrictions under which both levels and derivatives of individual demand functions as well as demands for non-exclusive public goods and most importantly the conditional sharing rule can be non-parametrically identified. In Section 1.4 we propose a non-parametric quantile regression procedure that closely follows this identification strategy and derive its asymptotic properties, which concludes the theoretical part of the paper. In Sections 1.5 and 1.6 we conduct a simulation study to investigate the finite sample behaviour of our estimation procedure before we estimate a simple collective labour supply model for Dutch households using the LISS panel.

Specification and Economic Restrictions

In this section, we will briefly introduce the collective model setting in its most general form [Chiappori & Ekeland, 2009] and present its key economic restrictions, which we will need to derive conditions for our identification strategy. While the model can be used for different purposes, e.g. labour supply and time-use settings, we will present the standard consumption-based version, without loss of generality. To capture heterogeneity at this stage, we index[1] households by $i \in I_N := \{1, \dots, N\}$ with members $s \in I_{S_i} := \{1, \dots, S_i\}$. While our approach would allow us to treat households of different sizes, we will without loss of generality assume that $S_i = S$. We assume that individual agents can choose from a finite number of consumable goods L, determining the set of alternatives $\mathbb{R}^L_+$, with corresponding prices $[p^0, p^s] \in \mathbb{R}^L_+$ and endowment $w \in \mathbb{R}$. We denote the number of publicly consumed goods by L_0 and the number of privately consumed goods by $L_1 = L - L_0$. For given prices and endowment, each household chooses $S + 1$ consumption vectors: assignable private consumption $(p^0, p^s, w) \mapsto x^s_i$ for each individual $s \in I_S$ in the household and public household consumption $(p^0, p^s, w) \mapsto x^0_i$, which cannot be assigned to any of the members. Then, for given prices $p := [p^1, \dots, p^S]$ and income w, we can define the set of feasible consumption bundles as:

$$\mathfrak{B}(p, p^0, w) := \left\{ (x^0, \dots, x^S) \in \mathbb{R}^{L_0 + (L - L_0)S}_+ : p^0 x^0 + \sum_{s \in I_S} p^s x^s \leqslant w \right\}. \tag{1.1}$$

This budget constraint set is convex, compact and contains the origin. We allow every individual to have their own set of preferences which we assume can be repre-

[1] We denote the index set running from 1 to J by $I_J := \{1, \dots, J\}$.

sented by a twice differentiable strictly quasi-concave utility function $u_i^s : \mathbb{R}_+^L \to \mathbb{R} : (x^0, x^s) \mapsto u_i^s$. Note that the structure of the utility function implies that there are no consumption externalities for private demands, since $x^{s'}$ does not enter u^s for $s \neq s'$. The collective choice of a household can then be viewed as an aggregate decision rule, depending on its member's utilities for a given consumption vector, paired with their relative bargaining 'power' within the household. This decision rule can be modelled by means of a social choice function $W_i : \mathbb{R}_+^S \to \mathbb{R}_+$ that assigns a positive real number representing the collective utility of the household to a vector of its members' individual utilities. Hence, from a household perspective, choosing an optimal consumption bundle from the feasible set of alternatives (1.1) is equivalent to finding the solution of the optimization problem

$$\max_{x \in \mathfrak{B}(p^s, p^0, w)} W_i(u_i^1(x), \ldots, u_i^S(x)). \tag{1.2}$$

As is common in the collective consumption literature, we only impose the following restriction on the bargaining structure of the model.

Assumption 1.1. [Intra-Household Pareto Efficiency] Let p, p^0 and w be given. If for any two vectors $(x_1^0, \ldots, x_1^S), (x_2^0, \ldots, x_2^S) \in \mathfrak{B}(p, p^0, w)$ it holds that $u^s(x_1^0, x_1^s) \geqslant u^s(x_2^0, x_2^s)$ for all $s \in I_S$, and $u^s(x_1^0, x_1^s) > u^s(x_2^0, x_2^s)$ for some $s \in I_S$, then the vector $(x_2^0, \ldots, x_2^S)$ is not a solution of the household decision process. The collection of remaining points shall be denoted as the Pareto frontier $\partial\mathfrak{B}(p^s, p^0, w)$, which is both non-empty and convex as it corresponds to the boundary of the budget set, by strict monotonicity of individual preferences (Walras' law).

It can be shown that any social welfare function W_i that satisfies Pareto efficiency, can be rewritten as a linear combination of individual utilities with weights $(p, w, z^\mu) \mapsto \mu_i^s$ corresponding to the individuals' *bargaining power*. These Pareto weights, are functions of prices, endowment and so-called *distribution factors* $z^\mu \in \mathbb{R}^{M_\mu}$ that enter μ but do not enter the individuals' preferences. We include the latter for the sake of completeness, however their existence is not necessary for our results, since we follow an identification strategy that requires observing intra-household allocation instead of distribution factors. Another consequence of Pareto efficiency is that the household optimization problem as described above can be written by means of an equivalent two-step procedure, in which the members agree on the sharing of total household income as well as consumption of public goods in the first stage and then, conditional upon this outcome, privately optimize individual consumption. In formal terms, this means that there exists a map $(p, p^0, w, z^\mu) \mapsto \rho_i^s$, the *conditional sharing rule*, such that every $[x, x^0](p, p^0, w, z^\mu)$ that solves (1.2) is also a solution of the following two-stage optimization process. In the first stage, using indirect individual utilities v_i^s from the second stage (below), the household optimizes

$$\max_{(x^0,\rho)\in\mathfrak{B}^0(p^0,w)} \sum_{s\in I_S} \mu_i^s \left(p^s, p^0, w, z^\mu\right) v_i^s \left(p^s, x^0, \rho^s\right) \tag{1.3}$$

with $\rho = [\rho^1, \dots, \rho^S]$ and budget constraint

$$\mathfrak{B}^0(p^0, w) := \left\{(x^0, \rho) \in \mathbb{R}^L \times \mathbb{R}^S : p^0 x^0 + \sum_{s\in I_S} \rho^s \leqslant w\right\}.$$

In the second stage, all members $s \in I_S$ of the household individually optimize, conditional upon optimal public consumption $x_i^0 = x_i^0(p, p^0, w, z^\mu)$ and the conditional sharing rule $\rho_i^s = \rho_i^s \left(p, p^0, w, z^\mu\right)$:

$$v_i^s(p^s, x_i^0(p, p^0, w, z^\mu), \rho_i^s(p, p^0, w, z^\mu)) = \max_{x^s\in\mathfrak{B}_i^s(p^s,p^0,w,z^\mu)} u_i^s \left(x^s, x_i^0(p, p^0, w, z^\mu)\right)$$

using their individual budget constraint

$$\mathfrak{B}_i^s \left(p^s, p^0, w, z^\mu\right) = \left\{x \in X : p^s x \leqslant \rho_i^s \left(p, p^0, w, z^\mu\right)\right\}.$$

The value function of the second stage v_i^s is called the *collective indirect utility of individual* (i, s) and is often of main interest. The second main building block, and arguably the most important feature of the collective model, is the *conditional sharing rule* ρ^s, on which we will focus in this paper. Note that here we follow an approach in which we assume that the nature of the good (private or public) is known a priori and private good consumption is assignable, such that we can allow members to simultaneously decide upon public good consumption and how the remaining endowment is allocated among the members, which is captured by the conditional sharing rule ρ^s for $s \in I_S$[2]. Thus, the individual shares are defined conditional upon of public good consumption x^0. However, if x^0 is invertible with respect to p^0 for given (p, w, z^μ), we can also express it in terms of (p, p^0, w, z^μ). A detailed discussion about this duality result can be found in [Chiappori & Ekeland, 2009].

[2]This stands in contrast with an alternative approach (see e.g. Browning et al. [2013]; Cherchye et al. [2015]) in which in the first stage the household members allocate their budget by choosing the respective shares. In the second stage individual consumption is then chosen subject to individualized Lindahl prices. This is equivalent to a model with only public goods with private goods being boundary cases for which Lindahl prices are equal to market prices.

In order to identify the latter, we have to link the structural economic model to observed behaviour. Since it is impossible to observe cardinalizations of direct utilities, it is a common strategy to focus on the solution of the utility maximization problem which takes the form of an observable demand system, and then use the underlying structure to derive restrictions these demands have to satisfy in order to be rationalizable. This is the problem of economic integrability, which has been studied for both the unitary [Hurwicz & Uzawa, 1971; Afriat, 1967] and the collective setting [Chiappori & Ekeland, 2006, 2009; Cherchye et al., 2007]. We will now briefly present the key economic restrictions of the collective model which we will use to provide necessary and sufficient conditions for identification of the conditional sharing rule, and refer the interested reader to the citations above for a more elaborate treatment of the topic.

For the first stage, using first order conditions we can determine public good consumption and the conditional sharing rule, which individuals decide upon given indirect utility from the second stage. For each household $i \in I_N$, a solution (x^0, ρ) of this first stage utility maximization problem (1.3) must satisfy for all $s \neq s'$

$$\mu_i^s(p, p^0, w, z^\mu)\frac{\partial v_i^s(p^s, x^0, \rho^s)}{\partial \rho^s} = \mu_i^{s'}(p, p^0, w, z^\mu)\frac{\partial v_i^{s'}(p^s, x^0, \rho^{s'})}{\partial \rho^{s'}} = m^0 \tag{1.4}$$

and additionally

$$\sum_{s \in I_s} \mu_i^s(p, p^0, w, z^\mu)\frac{\partial v_i^s(p^s, x^0, \rho^s)}{\partial x^0} = m^0 p^0, \tag{1.5}$$

where m^0 denotes the Lagrange multiplier arising from the budget constraint. Pareto efficiency implies that the budget constraint must hold with equality

$$p^0 x^0 + \sum_{s \in I_S} \rho^s = w.$$

As for the second stage, where each individual $(i, s) \in I_N \times I_S$ optimizes his own utility conditional upon the solution of the first stage, i.e. the conditional sharing rule and public consumption, solutions defining private consumption must solve the first order necessary and sufficient conditions

$$\nabla_{x^s} u_i^s\left(x_i^0(p, p^0, w, z^\mu), x^s\right) = m^s p \tag{1.6}$$

where m^s are Lagrange multiplies for each $s \in I_S$. According to Walras' law the solution lies on the boundary of the individual budget constraint such that $px^s = \rho_i^s\left(p, p^0, w, z^\mu\right)$, which is now defined in terms of the conditional sharing rule and can be interpreted as the share of endowment after public consumption (in monetary terms) that was allocated to member s in the previous stage.

Lemma 1. *The key economic restrictions implied by the first order conditions* (1.6) *of the second stage problem can be expressed as*[3]

$$\Xi_i^s(x^s) := \frac{\nabla_{x^s_{-L_1}} u_i^s}{\nabla_{x^s_{L_1}} u_i^s} - \frac{p^s_{-L_1}}{p^s_{L_1}} = 0\ . \tag{1.7}$$

Similarly, the first stage restrictions (1.4)-(1.5) *implicitly defining the optimal allocation of public goods and the conditional sharing rule can be rewritten as the* $(L_0 + S - 1)$*-dimensional nonlinear system of equations*

$$\Xi_i^0(x_0, \rho) := \boldsymbol{\Omega}_i(x_0, \rho)\boldsymbol{\mu}_i - \mathbf{c} := \begin{bmatrix} \nabla_{x^0}\boldsymbol{\nu}_i^T - \nabla_{\rho_1}\boldsymbol{\nu}_i^T \otimes p^0 \\ \left(\frac{\partial \nu_{i,1}}{\partial \rho_1}\right)^{-1} J_0 \nabla_\rho \boldsymbol{\nu}_i \end{bmatrix} \boldsymbol{\mu}_i - \begin{bmatrix} 0_{L_0} \\ \iota_{S-1} \end{bmatrix} = 0\ . \tag{1.8}$$

where $\boldsymbol{\nu}_i = [\nu_i^1, \ldots, \nu_i^s]^T$, $\boldsymbol{\rho}_i = [\rho_i^1, \ldots, \rho_i^s]^T$, $\boldsymbol{\mu}_i = [1, \mu_i^2/\mu_i^1, \ldots, \mu_i^S/\mu_i^1]^T$ *and* $(S-1) \times S$*-dimensional projection* $J_0 = \delta_{i,j+1}$ *with* δ *being defined as the Kronecker delta.*

Proof. See Appendix 1.A. □

Lemma 1 defines the first order conditions as systems of partial differential equations which characterize the economic restrictions imposed by the collective consumption model. If the functions x_i^s and $[x_i^0, \rho_i^s]$, are solutions to these systems for all members $s \in I_S$ of a particular household $i \in I_N$, we can conclude that they satisfy the restrictions imposed by the model and hence the household is rational in a collective sense. We can define a population to be rational if all individuals are collectively rational[4]. So far we have only considered representative consumers. The novelty of this paper will be the transition from this deterministic setting to a stochastic one in order to model unobserved heterogeneity. We will assume that all heterogeneity in individuals' preferences can be represented by a finite number of unobserved random variables and derive restrictions on how the latter may enter utility functions and Pareto weights by using the restrictions (1.7) and (1.8), which link these functions to the implicitly defined demand functions x^s, x^0 and conditional sharing rules ρ^s which are in turn functions of observed variables (p, p^0, z^μ, w) and the aforementioned unobserved variables which we will introduce in the next section.

[3]Subscript j refers to the j^{th} row of a vector, and $-j = \{j\}^c$ to all rows except j

[4]Note that, while this is not the primary purpose of this paper, in order to conduct welfare analysis one has to recover the underlying individual utility functions from observed data. Thus, certain integrability conditions known as the SNR(S-1) conditions need to be satisfied, which is a well established result in the literature [Chiappori & Ekeland, 2006, 2009].

Identification

Having discussed the economic restrictions of the collective model, in this section we will address the core question of this paper, namely whether and under which conditions, we can identify both private and public demands as well as the sharing rule[5] from observed variables. To model heterogeneous individuals we switch from the deterministic view of the model specified in Section 1.2 that applies to one individual $i \in I_N$ to a stochastic one, by introducing population utility functions and Pareto weights and express unobserved heterogeneity by a random variable entering these functions. To keep our results as general as possible we will neither specify the functional form of population utility functions and Pareto weights nor the distribution of the unobserved random variables, but rather treat the question of identification non-parametrically. It was already argued that the demand functions resulting from the two-stage optimization problem in Section 1.2, are in general non-linear and non-separable in the error terms, which is something we have to take into account. There exist at least two different strands in the literature considering identification of such functions, which both exploit information about the whole distribution of unobservables, instead of just its first or second moments as for example in Brown [1983] or Roehrig [1988], by using (local averages of) conditional quantiles. The first approach requires a function to be monotone Matzkin [2003], triangular [Chesher, 2003; Altonji & Matzkin, 2005; Imbens & Newey, 2009] or invertible Beckert & Blundell [2008] with respect to the unobserved error terms. A second route to achieve identification, which is more flexible with respect to excess heterogeneity, is the one proposed by Hoderlein & Mammen [2007, 2009]. While the latter allows random variables to be infinite-dimensional, it comes with the drawback of only applying to scalar functions or linear combinations of components of vector valued functions [Dette et al., 2016]. This is often sufficient for testing implications of rationality, however since we are dealing with a consumption setting with the inherent property of having a system of nonlinear equations which we want to estimate, we will make use of the first approach but will discuss a special case under which this approach can be applied. In what follows, we will provide sufficient conditions on both data availability and individual preferences, in form of non-parametric restrictions on the functional form of utility functions, that

[5]Having discussed different concepts of the (conditional) sharing rule in the previous section, we will from now on make the conditioning on public goods implicit by simply refering to this quantity as *the sharing rule*

ensure that all demand systems are invertible with respect to the unobserved error terms, exhibit no excess heterogeneity and hence meet the conditions required to be point identified.

As specified in Section 1.2, each household member $s \in I_S$ consumes $\left[x_i^s, x_i^0\right] \in \mathbb{R}_+^{L_1} \times \mathbb{R}_+^{L_0} = \mathbb{R}_+^L$ of which x_i^s is private and x_i^0 is public. Sharing is characterized by $\rho_i \in \mathbb{R}_+^S$, which sums up to total household income w_i.

Assumption 1.2. Individual consumption $\left(x_i^1, \dots, x_i^S\right)$, public consumption x_i^0 and hence the shares of household endowment $\rho_i = \left(\rho_i^1, \dots, \rho_i^S\right)$ are observed for all $i \in I_N$.

This assumption makes the data requirement for our proposed identification strategy explicit. Datasets providing the necessary information on intra-household allocation are becoming more and more popular, as for example the consumption module of the *Dutch LISS panel* [Cherchye et al., 2012], the *Danish Household Expenditure Survey* [Bonke & Browning, 2015], a survey of the *Italian International Center of Family studies* [Menon et al., 2012] and for time-use data the *UK Time Use Survey 2000* [Browning & Gortz, 2012]. In this general setting we assume there is in fact both public and individual private consumption $(L_0, L_1 > 0)$, which we believe is crucial to capture all behaviour that is inherent to a group of people living together as a family or household. However, there are two special cases which are both nested in our setting. The first one is a specification in which we only have public goods $(L = L_0)$. As one might guess, the second spacial case is one in which there is only private and assignable consumption $(L = L_1)$. In this case the observed sharing rule would immediately allow us to treat each individual separately in the context of a unitary consumption optimization setting.

To model heterogeneous individuals we define population utility functions $u^1, \dots, u^S$ and allow for individual specific taste-shifters defined as random vectors ε^0 and ε. Similarly we model unobserved heterogeneity with respect to intra-household bargaining by random distribution factors ε^μ entering the Pareto weights $\mu^1, \dots, \mu^S$. We could also allow for observed heterogeneity, taking e.g. demographic variables into account, but in the interest of readability we will abstract from this. In order to fully recover the primitives of our model, which allows us to perform welfare analysis on an individual level, we have to carefully balance the dimension of our random vectors. Thus, let ε^μ be a $(S-1)$-dimensional random vector that represents heterogeneity in the members

Pareto weights, which we will refer to as *bargaining shocks*. Further let $\varepsilon = [\varepsilon_1, \dots, \varepsilon_{L_1-1}]$ be a sequence of S-vectors capturing taste shocks of each individual for all privately consumed goods and let $\left[\varepsilon_1^0, \dots, \varepsilon_{L_0}^0\right]$ characterize the households taste shocks for the L_0 public goods.

Assumption 1.3. (i) Preferences can be represented by means of a utility function $u_i^s = u^s(x^s, x^0, \varepsilon^s, \varepsilon^0)$ which is twice continuously differentiable in all its arguments and for given $(\varepsilon^0, \varepsilon^s)$ strictly monotone and strictly quasi-concave in (x^s, x^0). Every member of a household is exposed to the same *taste shock* for a public good such that $\varepsilon_1^0, \dots, \varepsilon_{L_0}^0$ are all scalars and both $\nabla^2_{x^s,\varepsilon^s} u^s$ and $\nabla^2_{x^0,\varepsilon^0} u^s$ exist and have full rank for all $s \in I_S$.

(ii) Pareto weights can be represented as $\mu_i^s = \mu^s(p, w, z^\mu, \varepsilon_s^\mu)$, where μ^s continuously differentiable in (p, w, z^μ), strictly monotone in each element of z^μ and ε_s^μ, has range $(0,1)$ for all $s \in I_S$ and is normalized such that $\sum_{s \in I_S} \mu^s(p, w, z, \varepsilon_s^\mu) = 1$.

With this notion of taste- and bargaining-shocks, we can now define an individuals unobserved preferences and bargaining power by means of a random variable. While the general differentiability, and monotonicity assumptions on both utilities and Pareto weights are fairly standard in the literature, the assumption that members share a common taste shock for public goods and the rank conditions need further explanation. The former is needed, since we only observe, L_0 demands for public goods. If we would allow one taste shock for each public good and for each individual there would be excess heterogeneity in the first stage demand system and we would not be able to identify public good consumption and the sharing rule. The rank conditions ensure that taste shocks not only affect marginal utilities, but also that they enter in a non-ambiguous way, meaning that everything else equal, two different realizations of taste shocks do not induce the same marginal utilities. The one on $\nabla^2_{x^s,\varepsilon^s} u^s$ is well established in the literature for the unitary model [Beckert & Blundell, 2008] and $\nabla^2_{x^0,\varepsilon^0} u^s$ makes sure that the same holds for public goods and is hence a straightforward extension thereof.

To simplify notation throughout this section for all $s \in I_S$ we define $\pi^s = p^s$ and $\pi^0 = (p, p^0, w, z^\mu)$. Using Assumption 1.3 we can now write the first stage optimization problem, defined in (1.3), as

$$\begin{bmatrix} x^0 \\ \rho \end{bmatrix} \left(\pi^0, \pi^s, \varepsilon^0, \varepsilon^s, \varepsilon^\mu_s \ : s \in I_S\right) = \underset{(x^0,\rho)\in\mathbb{R}^{L_0+S-1}}{\arg\max} \sum_{s=1}^{S} \mu^s\left(\pi^s, \pi^0, \varepsilon^\mu_s\right) \nu^s\left(\pi^s, \rho^s, x^0, \varepsilon^s, \varepsilon^0\right)$$
$$\text{s.t. } p_i^{0T} x^0 + \sum_{s=1}^{S} \rho^s \leqslant w \tag{1.9}$$

It is easy to see, that in general the solution of this program depends on both public $\left[\varepsilon^0, \varepsilon^\mu\right]$ and private errors ε^s, whose dimensions are $L_0 + S - 1$ (Assumption 1.3.(i)) and $S(L_1-1)$ respectively, while the number of equations in the system $\left[x^0, \rho\right]$ is just L_0+S-1. Thus, we have excess heterogeneity in the sense that it is not possible to directly find a one-to-one mapping between a given realization of preference parameters $(\varepsilon, \varepsilon^0, \varepsilon^\mu)$ and observed demands and sharing rules (x^0, ρ). To overcome this, we will show under which conditions we can exploit information from demands for private goods x^s for all $s \in I_S$ to identify the distribution of ε which we denote as $\mathbf{P}$. It turns out that for given realization of ε a sufficient condition to achieve point identification of public demands and the sharing rule is given by

Assumption 1.4. (i) For all $s \in I_S$ and for all public goods x_l^0 where $l \in I_{L_0}$ it holds that

$$\frac{\partial^2 u^s}{\partial x_l^0 \partial \varepsilon_l^0} \geqslant \sum_{l' \in I_{L_0} \setminus l} \left| \frac{\partial^2 u^s}{\partial x_l^0 \partial \varepsilon_{l'}^0} \right|,$$

(ii) there *exists at least one* $s \in I_S$ such that

$$\frac{\partial^2 \nu^s}{\partial \rho^s \partial \varepsilon_l^0} = 0$$

(iii) and ε^0 is independent of ε for given π_0[6].

The first part of this assumption states that the effect of a taste shock for good l on the marginal utility on that good, must exceed the magnitude of aggregated cross-effects that taste shocks for all other goods have on this particular good l. The second part ensures that there exists at least one individual whose marginal utility with respect to income does not depend on the (common) taste shock for public goods.

Remember that Assumption 1.4 is only sufficient for identification for given private taste shocks ε^s or consistent predictions thereof. Since, the latter are characterized by

[6]Formally we define $((\varepsilon^0, \varepsilon^\mu), \varepsilon)$ as an element of the probability space $\left(\mathcal{E}^0 \times \mathcal{E}, \sigma(\mathcal{E}^0 \times \mathcal{E}), \mathbf{P}^{\varepsilon^0} \times \mathbf{P}\right)$ with $\mathcal{E}^0 = \mathbb{R}^{L_0+S-1}$, $\mathcal{E} = \mathbb{R}^{(L_1-1)S}$ and hence $\sigma(\mathcal{E}^0 \times \mathcal{E}) = \mathcal{B}(\mathbb{R}^{L_1 S+L_0-1})$ where the latter is the Borel σ-algebra defined on the respective Eucledian space. Without making it explicit in the notation, we allow the probability measures $\mathbf{P}$ and $\mathbf{P}^{\varepsilon^0}$ to depend on exogenous variables.

individual demands, identification of x^s for all $s \in I_S$ is crucial as well, for which we require one more structural assumption. We know that in the second stage all members $s \in I_S$ of the household conditionally upon optimal (ρ^s, x^0) optimize

$$x^s\left(\pi^s, \rho^s, x^0, \varepsilon^s, \varepsilon^0\right) = \underset{x \in \mathbb{R}^{L_1 - 1}}{\arg\max}\, u^s(x, x^0, \varepsilon^s, \varepsilon^0) \text{ s.t. } p^{s\mathsf{T}} x \leqslant \rho^s \ .$$

Using the solution of this optimization problem, which characterizes observed demands for private goods, we defined the resulting indirect utility function for each member as

$$v^s\left(\pi^s, \rho^s, x^0, \varepsilon^s, \varepsilon^0\right) = u^s\left(x\left(\pi^s, \rho^s, x^0, \varepsilon^s, \varepsilon^0\right), x^0, \varepsilon^s, \varepsilon^0\right)$$

which enters the first stage program defined in equation (1.9). While individual indirect utility functions are allowed to depend on both ε^s and $\left[\varepsilon^0, \varepsilon^\mu\right]$, in order to identify the $(L_1 - 1)$-dimensional private demand system, we need to ensure that $\left[\varepsilon^0, \varepsilon^\mu\right]$ do not enter the demand system x^s for any $s \in I_S$, which would cause excess heterogeneity.

Assumption 1.5. For each individual $s \in I_S$, the marginal rates of substitution between private goods does not depend on taste shocks for public goods

$$\nabla_{\varepsilon^0} \Xi^s(x^s, \varepsilon^s, \varepsilon^0) = 0.$$

Intuitively, this assumption states that after collectively choosing public consumption, the amount of private goods that is consumed by a member does not depend on the unobserved taste shock for the public good. A sufficient condition for this would be for example separability of the form $u^s(x^s, x^0, \varepsilon^s, \varepsilon^0) = G(g(x^s, \varepsilon^s), x^0, \varepsilon^0)$, for any two differentiable, increasing, real-valued functions G and g. This assumption is testable for observed x^0. To see this, note that the choice of x^s is conditional upon x^0 from the first stage. Hence x^s can be written as a function of either x^0 or π^0. Using prices and endowment poses restrictions on private demands, whereas using x^0 does not. Therefore under the null, i.e. if Assumption 1.5 holds, $x^s(\pi^s, x_i^0)$ should be "close to" $x^s(\pi^s, \pi^0)$.

Note that Assumptions 1.4 and 1.5 express sufficient conditions for the public demand system to be monotonic in the error terms. Necessary conditions are less economically traceable and therefore discussed only in the proof. Theorem 1 formalizes the main identification result of this section.

Theorem 1 (Identification). *Let* $\pi = [\pi^1, \ldots, \pi^S]$, $\Pi_\varepsilon^0 = [\pi^0, \pi, \varepsilon]$ *and* $\tau \in (0,1)$. *Under assumptions 1.1-1.5, both levels* $Q^\tau_{x_j^s|\pi^s}(\pi^s)$ *and derivatives* $\dot{Q}^\tau_{x_j^s|\pi^s}(\pi^s)$ *of the* τ^{th} *conditional* *quantile*[7] *of each component* $j \in I_{L_1-1}$ *of private demands are identified. In addition to this, for given individual private taste shocks* ε*, levels* $Q^{\tau}_{(x^0_l,\rho)|\Pi^0_\varepsilon}(\Pi^0_\varepsilon)$ *and derivatives* $\dot{Q}^{\tau}_{(x^0,\rho)_l|\Pi^0_\varepsilon}(\Pi^0_\varepsilon)$ *of each component* $l \in I_{L_0+S-1}$ *of public demands and the sharing rule are also identified.*

Proof. In order to understand the mechanics of the underlying identification strategy it is instructional to look at the main steps of the proof. Note that under Assumptions 1.2 and 1.5, the second stage private demand problem is equivalent to the unitary model, for which invertibility is well established [Beckert & Blundell, 2008]. We will therefore only focus on the first stage sharing rule and public goods at this point. For the complete and detailed version of the proof we refer the interested reader to Appendix 1.A.

The left hand side of our first order conditions defined in (1.8) which implicitly defines $\left[x^0, \rho\right]$ can be written as

$$\Xi^0\left(\left[x^0, \rho\right], \Pi^0_\varepsilon, \left[\varepsilon^0, \varepsilon^\mu\right]\right) = \Omega(x^0, \boldsymbol{\rho}, \Pi^0_\varepsilon, \varepsilon^0)\boldsymbol{\mu}(\Pi^0_\varepsilon, \varepsilon^\mu) - \mathbf{c}$$

where we define Ω_1 and Ω_2 to be the first L_0 and the remaining $S-1$ rows of the matrix, respectively. Intuitively Ω_1 largely refers to the restrictions defining public good consumption, whereas Ω_2 defines sharing rules in terms of Pareto weights.

We can now locally write $\left[x^0, \rho\right]$ in terms of Π^0_ε and $\left[\varepsilon^0, \varepsilon^\mu\right]$ using the implicit function theorem for vector valued functions

$$\begin{aligned}\nabla_{[\varepsilon^0,\varepsilon^\mu]}\left[x^0, \rho\right]\left(\Pi^0_\varepsilon, \left[\varepsilon^0, \varepsilon^\mu\right]\right) = &-\left[\nabla_{[x^0,\rho]}\Xi^0\left(\left[x^0, \rho\right], \Pi^0_\varepsilon, \left[\varepsilon^0, \varepsilon^\mu\right]\right)\right]^{-1} \\ &\times \nabla_{[\varepsilon^0,\varepsilon^\mu]}\Xi^0\left(\left[x^0, \rho\right], \Pi^0_\varepsilon, \left[\varepsilon^0, \varepsilon^\mu\right]\right).\end{aligned}$$

Invertibility with respect to $\left[\varepsilon^0, \varepsilon^\mu\right]$ requires this matrix to be of full rank (L_0+S-1). Note that the inverse of $\nabla_{[x^0,\rho]}\Xi^0$ exists by construction of the collective model. Hence it remains to show that $\nabla_{[\varepsilon^0,\varepsilon^\mu]}\Xi^0$ has full rank. Writing the latter as a block matrix

$$\Psi = \nabla_{[\varepsilon^0,\varepsilon^\mu]}\Xi^0 = \left[\nabla_{\varepsilon^0}\Xi^0\ ,\ \nabla_{\varepsilon^\mu}\Xi^0\right] = \begin{bmatrix}\Psi_{11} & \Psi_{12} \\ \Psi_{21} & \Psi_{22}\end{bmatrix}$$

with

$$\Psi_{j\bullet} := \nabla_{[\varepsilon^0,\varepsilon^\mu]}\Xi^0 = \left[\nabla_{\varepsilon^0}\Xi^0\ ,\ \nabla_{\varepsilon^\mu}\Xi^0\right] = \left[\left(\mu^\top \otimes I_{L_0+S-1}\right)\nabla_{\varepsilon^0}\mathrm{vec}\Omega_j\ ,\ \Omega_j\nabla_{\varepsilon^\mu}\boldsymbol{\mu}\right]$$

we can now analyze the respective blocks separately. Intuitively, since Ψ_{11} refers to the impact of taste shocks for public goods on the equations defining public goods

[7] The quantile of a r.v. Y conditional on X, with c.d.f. $F_{Y|X}$ is defined as $Q^\tau_y(x) := \inf\left\{y \in \mathbb{R} : F_{Y|X=x}(y) \geqslant \tau\right\}$.

and Ψ_{22} refers to the impact of bargaining shocks on the sharing rule, we allow (and require) these matrices to be largely unrestricted and have full rank. Ψ_{12} and Ψ_{21} on the other define the impact of bargaining shocks on public good consumption and that of taste shocks on the sharing rule, respectively. We will show that the latter is zero, which follows from the fact that ratios of Pareto weights are inverse ratios of individual marginal utilities with respect to the sharing rule as defined in equation (1.4) and the fact that by Assumption 1.3.(ii) the former do not depend on public taste shocks. It than follows from taking the Schur complement Ψ/Ψ_{22} that $\text{rk}\,(\Psi) = \text{rk}\,(\Psi/\Psi_{22}) + \text{rk}\,(\Psi_{22}) = \text{rk}\,(\Psi_{11}) + \text{rk}\,(\Psi_{22})$ since $\Psi/\Psi_{22} = \Psi/\Psi_{11} + \Psi_{12}\Psi_{22}^{-1}\Psi_{21}$ where the latter term is zero. Ψ_{11} can be written as a linear combination of second derivatives $\nabla^2_{x^0,\varepsilon^0}\nu^s$. Assumption 1.3.(i) ensures that each of them has full rank whereas Assumption 1.4 ensures positive semi-definiteness of each of them such that the linear combination has full rank too. Assumption 1.3.(ii) immediately implies full rank of Ψ_{22} since $\nabla_{\varepsilon^\mu}\mu$ is diagonal with all elements being non-zero. Thus, $\text{rk}\,(\Psi) = \dim\Psi = L_0 + S - 1$ and public demands and the sharing rule can be locally inverted with respect to $\left[\varepsilon^0, \varepsilon^\mu\right]$ and are thus identified. □

It is important to emphasize that the second identification result regarding public demands and the sharing rule requires identification of the distribution $\mathbf{P}$ of private taste shocks ε which are unobserved. While, we have shown invertibility of the systems with respect to error terms which is both necessary and sufficient for identification of conditional quantiles of these demands, it is not sufficient (however necessary) for the identification of $\mathbf{P}$. In this more general setting, all we can identify is the distribution of a generalized error term for each $s \in I_S : \alpha(\pi^s) = x - Q^\tau_{x|\pi^s}(\pi^s)$. Without further restrictions there is no one-to-one mapping between α and ε. This is a known problem within the identification literature and hence the same logic applies to the conditions which Beckert & Blundell [2008] provide for the unitary model. For this reason, we will now provide examples and refinements under which we can identify not only the distribution of α but also the one of ε.

Collective labour supply model

This model refers to a collective model with two private goods $L_1 = 2$. Labour supply is modelled as consumption of leisure and one composite consumption good [Chiappori, 1992; Fortin & Lacroix, 1997; Blundell et al., 2007; Cherchye et al., 2015]. Using the budget constraint, the observed demand system reduces to a scalar leisure demand function. There is no restriction on the number of public goods L_0, due to the observability of the sharing rule. Invertibility as discussed above for such a function implies monotonicity with respect to the error term such that, using the invariance property of conditional quantiles to monotone transforms, the distribution of ε^s is identified for each $s \in I_S$ [Matzkin, 2003]. In fact this will be the model setting that we will employ in our empirical application in Section 1.6.

General consumption model with triangularity in private demands

This refinement is a generalization of the aforementioned monotonicity approach and follows the lines of Chesher [2003] and Imbens & Newey [2009]. If, in addition to the invertibility restrictions we derived we also have triangularity with respect to the error terms, which means that we can order goods with respect to their error structure where $x^s_{L_1}$ is only affected by $\varepsilon^s_{L_1}$, $x^s_{L_1-1}$ is affected by $\varepsilon^s_{L_1}$ and $\varepsilon^s_{L_1-1}$ and so on, we can also identify the distribution of each ε^s for all $s \in I_S$. An example for a parametrization that satisfies this triangularity condition is the data generating process we specify in our consumption model within the Monte Carlo study in Section 1.5.

The case of two-person households without public goods

This case is somewhat different from the previous two cases, as it does not require estimation of private taste shocks at all. Assume that all goods are consumed privately ($L_0 = 0$) and the group consists of only two members ($S = 2$), as for example a typical household without children. In such a setting, excess heterogeneity is not an issue since it is well established [Hoderlein & Mammen, 2007, 2009] that a local polynomial quantile estimator, as we will propose it in the next section, can in fact be used to estimate local averages of conditional quantiles and its derivatives of such function. Since this approach allows for a very general (infinite-dimensional) error structure, taste shocks for public goods and bargaining shocks can be unrestricted such that one can drop assumption 1.4. In addition to this, one could allow for member specific taste shocks for public goods instead of common ones.

Corollary 1.1. If for given π^s there exists a one-to-one mapping between ε and generalized errors $a(\pi^s) = x - Q^\tau_{x|\pi^s}(\pi^s)$, then identification of $\mathbf{P}_a$ is sufficient for identification of both level $\mathbb{E}Q^\tau_{(x^0_l,\rho)|(\pi^0,\pi)}(\pi^0,\pi) := \int_\mathcal{E} Q^\tau_{(x^0_l,\rho)|\Pi^0_\varepsilon}(\Pi^0_\varepsilon)$ and derivative $\mathbb{E}\dot{Q}^\tau_{(x^0,\rho)|(\pi^0,\pi)}(\pi^0,\pi) := \int_\mathcal{E} \dot{Q}^\tau_{(x^0,\rho)|\Pi^0_\varepsilon}(\Pi^0_\varepsilon)d\mathbf{P}$.

Proof. This is an immediate consequence of Matzkin [2003] for monotone functions (first refinement) and Chesher [2003] for triangular systems (second refinement). □

In the next section we will study how we can consistently estimate the distribution of individual private taste shocks using private demands, as well as local averages of public goods and the sharing rule as defined in Corollary 1.1.

Estimation of Demands and the Sharing Rule

Information on intra-household allocation of goods allows us to follow an empirical strategy that estimates the sharing rule and demands for public goods separately from the individuals' demand functions for private goods. However, we have seen that an individual's taste-shocks ε^s for private goods, will in general also influence the household decision about the sharing rule and public good consumption. This is due to the fact that they are characterized by a utility maximization problem, that is a combination of the members' utilities. Taste shocks are unobserved and we do not want to specify their distribution *a priori*. Nevertheless, we can approximate them using the residuals from the private demand estimation. Once we obtain such predictions, we are able to estimate the demands for public goods and the sharing rule conditional upon them. Finally, since the effect of private taste shocks has no economic meaning per se, we are going to take expectations with respect to them in order to obtain a local average of a conditional quantile.

We have shown in Theorem 1 and Corollary 1.1 under which restrictions conditional quantiles of x^s_j, $[\rho^1, \dots, \rho^{S-1}, x^0]_l$ are identified. For notational simplicity we will omit superscript $s \in I_S$ and subscripts $j \in I_{L_1-1}$ and $l \in I_{L_0+S-1}$ from now on and write only ρ instead of $(\rho, x^0)_l$ and x instead of x^s_j to refer to the respective components. We define the dimension of the exogenous variables by at $K_0 = L_0 + SL_1 + 1$ and $K = L_1$ for ρ and x, respectively. Before we present the estimation procedure, which is closely related to our identification results, we need a weak regularity assumption regarding the distribution of the data and unobserved variables.

Assumption 1.6. (i) $(x, \pi)_i$ and $(\rho, \pi^0)_i$ are i.i.d. sequences and have non-degenerate distribution functions with Lipschitz-continuous conditional densities $f_{x|\pi}$, $f_{\rho|\pi^0}$ and marginal densities f_π, f_{π^0}, respectively, which are all uniformly bounded by a finite constant $M < \infty$ and non-zero at $F^{-1}_{x|\pi}(\tau)$ and $F^{-1}_{\rho|\pi_0}(\tau)$, respectively.

(ii) ε_i and ε_i^0 are i.i.d., have expectation zero, finite second moments and are characterized by their continuous densities f_ε and f_{ε^0}.

(iii) ε_i and ε_i^0 are independent of π_i and π_i^0 respectively

(iv) $\pi \mapsto Q^\tau_{x|\pi}$ and $\pi_0 \mapsto Q^\tau_{\rho|\pi_0}$ have finite, continuous second derivatives denoted by $\ddot{Q}^\tau_{x|\pi}$ and $\ddot{Q}^\tau_{\rho|\pi_0}$, respectively.

To obtain estimates for the conditional quantile [Koenker & Bassett, 1978] of a private demand x and its corresponding gradient vector, in a first stage[8] we use a local linear approximation [Chaudhuri, 1991; Yu & Jones, 1998] of x around some π_0 and minimize

$$\left(\widehat{Q^\tau_{x|\pi}}(\pi_0), \widehat{\dot{Q}^\tau_{x|\pi}}(\pi_0)\right) = \arg\min_{\gamma_0,\gamma_1} \sum_{i=1}^N \rho_\tau\left(x_i - \gamma_0 - \gamma_1^T(\pi_i - \pi_0)\right) K\left(\frac{\pi_i - \pi_0}{h}\right)$$

where ρ_τ is the quantile loss (check-)function $u \mapsto u(\tau - \mathbb{1}_{\{u>0\}})$, K is a smooth, symmetric kernel function with compact support $[a, b]$ and variance one, that puts decreasing weight on observations far from π_0 and h a free bandwidth parameter tending to zero as $N \to \infty$. Since this objective function is not differentiable, we do not have an explicit solution for our quantities of interest unlike for the local linear mean regression. Although it constitutes a fairly standard result in the literature, for reasons of self sufficiency of the paper and the fact that it enters the second stage estimates, we provide the asymptotic distribution for this first stage estimate in Lemma 2.

Lemma 2 (First stage asymptotic distribution). *Let* $K(\cdot)$ *be a symmetric Kernel with bounded support and finite first derivative* $\dot{K}(\cdot)$*. Under Assumptions 1.1-1.6, as* $h \to 0$ *and* $H_N := Nh^{L_1} \to \infty$*:*

$$\sqrt{H_N}\begin{pmatrix} \widehat{Q^\tau_{x|\pi}}(\pi_0) - Q^\tau_{x|\pi}(\pi_0) \\ h(\widehat{\dot{Q}^\tau_{x|\pi}}(\pi_0) - \dot{Q}^\tau_{x|\pi}(\pi_0)) \end{pmatrix} \rightsquigarrow \mathcal{N}\left(\frac{h^2}{2}B_0(\pi_0), \frac{\tau(1-\tau)B_1}{f_\pi(\pi_0) f^2_{x|\pi}(Q^\tau_{x|\pi}(\pi_0))}\right)$$

with

$$B_{0j}(\pi_0) = tr\left\{\ddot{Q}^\tau_{x|\pi}(\pi_0)\int uu^T \begin{pmatrix}1\\u\end{pmatrix}_j K(u)\,du\right\}$$

and

$$B_1 = \begin{bmatrix} \int K^2(u)du & 0 \\ 0 & \int uu^T K^2(u)du \end{bmatrix}.$$

Proof. See Appendix 1.A. □

[8]Note that first and second stages are now referring to the estimation steps and are not to be confused with the ones that were defined in Section 1.2.

As is common in local linear regression, we get a bias of second order. The variance is determined by τ, the selected kernel, the marginal density of the independent variables and the sparsity function (i.e. the inverse conditional density function evaluated at the true quantile) which can both be estimated given our observed data. It is noteworthy that the function and its gradient vector are asymptotically independent which follows from the block-diagonal nature of B_1. This result stems from the fact that we use a symmetric kernel. Defining $\widehat{a}(\pi_i) = x_i - \widehat{Q^{\tau}_{x|\pi}}(\pi_i)$ as the generalized residual that arises from the non-separable nature of the function we estimate, we get $\widehat{Q^{\tau}_{x|\pi}}(\pi_i) - Q^{\tau}_{x|\pi}(\pi_i) = \widehat{a}(\pi_i) - a(\pi_i)$. These residuals represent estimates for the generalized error $a(\pi_i) = x_i - Q^{\tau}_{x|\pi}(\pi_i)$, which constitutes a monotone function of the underlying taste shock ε and can hence be used as an approximation of the latter under the refinements provided in the previous section. Since $\widehat{a}(\pi_0)$ is defined by $\widehat{\gamma}(\pi_0) = \widehat{Q^{\tau}_{x|\pi}(\pi_0)}$ the dependence is made explicit by writing $\pi^0_0 = \pi^0_0(\widehat{\gamma})$. This allows us to define the second stage objective function in terms of the estimates of the first stage:

$$\left(\widehat{Q^{\tau}_{\rho|\pi^0}}(\pi^0_0), \widehat{\dot{Q}^{\tau}_{\rho|\pi^0}}(\pi^0_0)\right) = \arg\min_{\theta_0,\theta_1} \sum_{i=1}^{N} \rho_\tau \left(\rho_i - \theta_0 - \theta_1^T(\pi^0_i(\widehat{\gamma}) - \pi^0_0)\right) K\left(\frac{\pi^0_i(\widehat{\gamma}) - \pi^0_0}{h_0}\right)$$

Paired with the non-differentiable nature of this second stage objective function, the dependence on the first stage parameters makes our analysis slightly more complicated. In order to show that the minimum with respect to θ is obtained uniformly with respect to the value of γ, we employ empirical process techniques to proof consistency of the second stage estimates. We omit the details here and refer the interested reader to Lemma 6 (*stochastic equicontinuity*) and Lemma 7 (*consistency*) in Appendix 1.A.

In order to obtain an asymptotic distribution for the second stage estimates we use a Bahadur representation stated in Lemma 3 which, as one might expect, depends on the approximation error of the first stage estimates.

Lemma 3 (Bahadur representation 2$^{\text{nd}}$ stage)**.** *Let* $K(\cdot)$ *be a symmetric Kernel with bounded support and finite first derivative* $\dot{K}(\cdot)$*. Under Assumptions 1.1-1.6, as* $h, h_0 \to 0$ *and* $H_N, H_{0,N} \to \infty$*:*

$$\sqrt{H_{0,N}}\begin{pmatrix}\widehat{Q^{\tau}_{\rho|\pi^0}}(\pi^0_0)-Q^{\tau}_{\rho|\pi^0}(\pi^0_0)\\ h_0(\widehat{\dot{Q}^{\tau}_{\rho|\pi^0}}(\pi^0_0)-\dot{Q}^{\tau}_{\rho|\pi^0}(\pi^0_0))\end{pmatrix}=$$
$$-\Gamma^{-1}_{\theta,0}\frac{1}{\sqrt{H_{0,N}}}\sum_{i\in I_N}\left(\tau-\mathbb{1}\left\{\rho_i\leqslant\theta^{\mathsf{T}}z^0_i(\gamma)\right\}\right)\begin{pmatrix}1\\ \frac{\pi^0_i-\pi^0_0}{h_0}\end{pmatrix}K\left(\frac{\pi^0_i-\pi^0_0}{h_0}\right)$$
$$+\sqrt{\frac{H_{0,N}}{H_N}}\Gamma^{-1}_{\theta,0}\Gamma_{\gamma,0}D_0^{-1}\frac{1}{\sqrt{H_N}}\sum_{i\in I_N}\left(\tau-\mathbb{1}\left\{x_i\leqslant\gamma^{\mathsf{T}}z_i\right\}\right)\begin{pmatrix}1\\ \frac{\pi_i-\pi_0}{h}\end{pmatrix}K\left(\frac{\pi_i-\pi_0}{h}\right)+o_P(1)$$

Proof. See Appendix 1.A. □

First and second stage estimators converge at non-parametric rates $\sqrt{H_N}$ and $\sqrt{H_{0,N}}$ respectively. Due to the higher dimension of π^0 compared to π^s by construction (K_0 and K respectively) and the fact that they both have finite variances, letting $c_0, c_1, c_2 > 0$ and using the optimal bandwidths we get $\sqrt{H_{0,N}}=\sqrt{Nh^{K_0}_{0,n}}=N^{\frac{1}{2}-\frac{K_0}{2(K_0+4)}}c_0$ and $\sqrt{H_N}=\sqrt{Nh^K_n}=N^{\frac{1}{2}-\frac{K}{2(K+4)}}c_1$ such that the term $\sqrt{\frac{H_{0,N}}{H_N}}=c_2N^{\frac{4(K-K_0)}{2(K_0+4)(K+4)}}$ converges to zero and the Bahadur representation converges to a (L_0+SL_1-1)-dimensional Brownian bridge.

Since our main interest lies not in the conditional quantile itself, but rather in its local average with respect to private taste shocks, in a final step we will integrate over our estimates $\widehat{Q^{\tau}_{\rho|\pi^0}}(\pi^0_0,a)$ with respect the distribution of a. However, as we do not know their distribution F_a, but only the empirical distribution of $\widehat{a}$ which we denote as $\widehat{F}_{\widehat{a}}$, some further work needs to be done. In order to be able to draw n realizations from the empirical distribution function $\widehat{F}_{\widehat{a}}$ instead from the unknown true one F_a and then take the sample average, we have to first show that the corresponding law $\sqrt{n}(\widehat{\mathbb{P}}_n-\mathbf{P}_n)$ converges to the law $\sqrt{n}(\mathbb{P}_n-\mathbf{P}_0)$. This uniformity result with respect to the underlying measure follows from van der Vaart & Wellner [1996, Theorem 2.8.9] and Lemma 8 which is again omitted here. Consistency and the asymptotic distribution of the numerically integrated second stage estimator which estimates the local average conditional quantile of the sharing rule are derived in Theorem 2.

Theorem 2 (Asymptotic distribution of local average conditional quantile)**.** *Let* $a^*_i\sim\widehat{F}_{\widehat{a}}$ *for* $i\in I_n$ *and* $H_{0,N}=Nh_0^{L_0+SL_1+1}$ *where* $\mathcal{O}(H_N)=o(n)$*. Then under Assumptions 1.1-1.6, as* $h_0\to 0$ *and* $H_{0,N}\to\infty$*,*

$$\sqrt{H_{0,N}}\begin{pmatrix}n^{-1}\sum_{i\in I_n}\widehat{Q^{\tau}_{\rho|\pi^0}}(\pi^0_0,a^*_i)-\int Q^{\tau}_{\rho|\pi^0}(\pi^0_0,a)dF_a(a)\\ n^{-1}\sum_{i\in I_n}h_0(\widehat{\dot{Q}^{\tau}_{\rho|\pi^0}}(\pi^0_0,a^*_i)-\int\dot{Q}^{\tau}_{\rho|\pi^0}(\pi^0_0,a)dF_a(a))\end{pmatrix}\rightsquigarrow\mathcal{N}\left(\mathbb{B}(\pi^0_0),\mathbb{V}(\pi^0_0)\right)$$

with $\mathbb{B}(\pi_0^o) = \frac{h_0^2}{2} B_0^o \int \ddot{Q}^{\tau}_{\rho|\pi^o}\left(\pi_0^o(a)\right) dF_a(a)$ *and* $\mathbb{V}(\pi_0^o)$ *defined in equations* (1.24) *and* (1.25) *at the end the proof.*

Proof. See Appendix 1.A. □

Theorem 2 constitutes the second main contribution of this paper, the asymptotic properties of our proposed estimation procedure. It can be seen that, although there is again a second order bias, we can consistently estimate the covariance matrix of the local average conditional quantile.

To sum up, if there exists information on intra-household allocation of consumption and we are willing to impose the structural assumptions provided in the previous section regarding the behaviour of preferences with respect to taste shocks, it is possible to non-parametrically estimate the conditional sharing rule and demands for both private and public consumption goods.

Monte Carlo Simulations

The purpose of this section is two-fold. Not only will we study the finite sample behaviour of the estimator proposed in the previous section, the specification of a particular structure will also serve to give some insight in how the model is solved and how our identification strategy works.

As is common in the demand estimation literature we will start by specifying a data-generating process in terms of a conditional indirect utility function constituting individual preferences as a function of prices for private goods, the individuals' share and public consumption. For this, we use the following indirect utility function with two separable sub-utilities for private and public goods:

$$v^s(p^s, \rho^s, x^0, \varepsilon^s, \varepsilon^0) = \frac{\log \rho^s - \log a_s(p^s)}{b_s(p^s, \varepsilon^s)} + \eta_s(\varepsilon^0) \log x^0$$

with associated price indices

$$\log a_s(p^s) = \alpha_s^T [1, \tilde{p}^s] + \tilde{p}^{sT} \Gamma_s \tilde{p}^s$$
$$b_s(p^s, \varepsilon^s) = \prod_{l \in I_{L_1}} (p_l^s)^{\beta_l^s(\varepsilon_l^s)}$$

where $\tilde{p}^s = [\log p_1^s, \ldots, \log p_{L_1}^s]$ for all $s \in I_S = \{1, 2\}$, and the amount of private goods is $L_1 = 3$. The first term is the indirect utility function that will generate an *almost ideal demand system* [Deaton & Muellbauer, 1980]. The second term constitutes a sub-utility function characterizing preferences for the public good, represented by means of a Cobb-Douglas utility function, in this case with only one public good. Note that these terms need not be additive; any sufficiently separable function satisfying Assumption 1.5 can be specified. Unobserved heterogeneity with respect to good preferences is modeled using random coefficients $\beta_s(\varepsilon^s) = \beta_s + \varepsilon^s$ for private taste shocks and $\eta_s(\varepsilon^0) = \eta_s + \varepsilon^0$ for public taste shocks. Remember that we must not have excess heterogeneity, i.e. the length of the vector ε^s cannot exceed the number of freely chosen private goods $L_1 - 1 = 2$ for any $s \in I_S$. In addition to this, public taste shocks ε^0 have to be common among spouses, according to Assumption 1.3(i). Note that in this specification the last assumption is not restrictive since the random coefficient $\eta_s(\varepsilon^0)$ is linear in the error term. Hence, once we take linear combinations of the individuals' indirect utility functions which has an additive sub-utility for the public good, the household taste-shocks ε^0 can be interpreted as a linear combination of individual taste shocks with respect to public goods with weight determined by the individuals' bargaining power. The second main ingredient of the collective model is the aggregation rule. Pareto-efficient social welfare functions can be written as linear combinations of individual utilities. For our simulations, we will follow the convention and specify Pareto weights as the logistic function with an index that is a linear combination of prices, distribution factors and unobserved heterogeneity in bargaining as an additive error:

$$\mu^s(w, p, z^\mu_s, \varepsilon^\mu_s) = \left(1 + \exp(-(\gamma_{s,0} + \gamma^T_{s,1} p + \gamma^T_{s,2} z^\mu_s + \varepsilon^\mu_s))\right)^{-1}.$$

It should again be emphasized at this stage that the existence of a distribution factor is not required for our identification strategy. This gives us the first stage problem:

$$\max_{x^0,\rho^1,\rho^2} \nu^1(p^1, \rho^1, x^0, \varepsilon^1, \varepsilon^0)\mu(w, p, z^\mu, \varepsilon^\mu) + \nu^2(p^2, \rho^2, x^0, \varepsilon^2, \varepsilon^0)\,(1 - \mu(w, p, z^\mu, \varepsilon^\mu))$$
$$\text{s.t. } \rho^1 + \rho^2 + p^0 x^0 \leqslant w,$$

where we let $p = (p^1, p^2)$ and drop the subscript s for common or restricted variables.

Lemma 4 (Almost Ideal Demand System)**.** *Demands for public goods and the sharing rule are defined by*

$$\begin{pmatrix} x_0 \\ \rho \end{pmatrix} = \frac{w}{p^0} \frac{b_2(p^2, \varepsilon^2)}{\overline{b}(p^1, p^2, w, z^\mu, \varepsilon^1, \varepsilon^2, \varepsilon^\mu) + b_1(p^1, \varepsilon^1) b_2(p^2, \varepsilon^2)\overline{\eta}(p^1, p^2, w, z^\mu, \varepsilon^0, \varepsilon^\mu)} \times \begin{pmatrix} b_1(p^1, \varepsilon^1)\overline{\eta}(p^1, p^2, w, z^\mu, \varepsilon^0, \varepsilon^\mu) \\ p^0 \mu(p^1, p^2, w, z^\mu, \varepsilon^\mu) \end{pmatrix} \tag{1.10}$$

which are functions of both private and public taste shocks where $\overline{\eta} = \mu\eta_1 + (1-\mu)\eta_2 + \varepsilon^0$ *and* $\overline{b} = \mu b_1(\varepsilon^1) + (1-\mu) b_2(\varepsilon^2)$.

This constitutes a non-linear system of $L_0 + S - 1 = 2$ equations with $K^0 = 9$ exogenous variables $(p^0, p_1^1, p_2^1, p_3^1, p_1^2, p_2^2, p_3^2, z^\mu, w)$ and the predictions for the 4 private taste shocks $(\varepsilon_1^1, \varepsilon_2^1, \varepsilon_1^2, \varepsilon_2^2)$. As for private goods, since their demands are characterized by a standard individual utility optimization problem with indirect utility v^s, it is straightforward to solve for *Walrasian demands* x^s using *Roy's identity* to get

$$x^s = \rho^s \tilde{\tilde{p}}^s \left(\alpha^s + \Gamma^s \tilde{p^s} + \beta_s(\varepsilon^s) \log \left(\frac{\rho_s}{a_s(p^s)} \right) \right)$$

where $\tilde{\tilde{p}}^s = \mathrm{diag}(1/p_1^s, \ldots, 1/p_{L_1}^s)$, resulting in a system of $L_1 - 1 = 2$ equations with $K = 4$ exogenous variables $(p_1^s, p_2^s, p_3^s, \rho^s)$ for each $s \in I_S$. It can be seen that this system can be inverted with respect to ε^s and each component is a monotone function in the error term, such that the data generating process satisfies our triangularity refinement.

To finalize the specification of the data-generating process, we have to associate numerical values with the model parameters. Note that we are not after identifying these parameter values themselves, but only functions thereof, namely private and public demands as well as the sharing rule. We pick the parameter values that define the price indices of the almost ideal demand system a_s and b_s in a way, such that the properties of symmetry, homogeneity of degree zero and adding-up are imposed:

$$\alpha_1 = \alpha_2 = \begin{pmatrix} 0 \\ .3 \\ .5 \\ .2 \end{pmatrix}, \beta_1 = \begin{pmatrix} .04 \\ -.06 \\ -.1 \end{pmatrix}, \beta_2 = \begin{pmatrix} .06 \\ -.02 \\ -.04 \end{pmatrix},$$

$$\Gamma_1 = \begin{pmatrix} 1. & -.3 & -.7 \\ -.3 & .8 & -.5 \\ -.7 & -.5 & 1.2 \end{pmatrix}, \Gamma_2 = \begin{pmatrix} .8 & -.4 & -.4 \\ -.4 & 1.1 & -.7 \\ -.4 & -.7 & 1.1 \end{pmatrix}.$$

The parameters $\eta_1 = 0.05$ and $\eta_2 = .15$, define the relative importance of the public good for each individual. In this case, the wife $(s = 2)$ values the public good more than the husband $(s = 1)$. As for the aggregation rule, keeping things simple we only define them as a function of distribution factors by picking $\gamma_{1,0} = \gamma_{1,1} = 0$ and $\gamma_{1,2} = 1$.

Exogenous variables are generated independently of each other according to

$$p^1, p^2 \sim U[17{,}23]^{L_1}, \quad p^0 \sim U[8{,}12]^{L_0}, \quad w \sim U[700{,}1300], \quad z^\mu \sim U[-2{,}2]$$

and finally unobserved variables are drawn from the distributions

$$\varepsilon^0 \sim U[-.25, .25], \quad \varepsilon^\mu \sim U[-1{,}1], \quad \varepsilon^1, \varepsilon^2 \sim \mathcal{N}\left(\begin{pmatrix} 0 \\ 0 \end{pmatrix}, \begin{pmatrix} 2.5 \times 10^{-4} & 1 \times 10^{-4} \\ 1 \times 10^{-4} & 2.5 \times 10^{-4} \end{pmatrix} \right).$$

For simplicity and without loss of generality we choose not to include observed heterogeneity. For each realization we then calculate public good consumption and the sharing rule, which is then used to calculate private consumption.

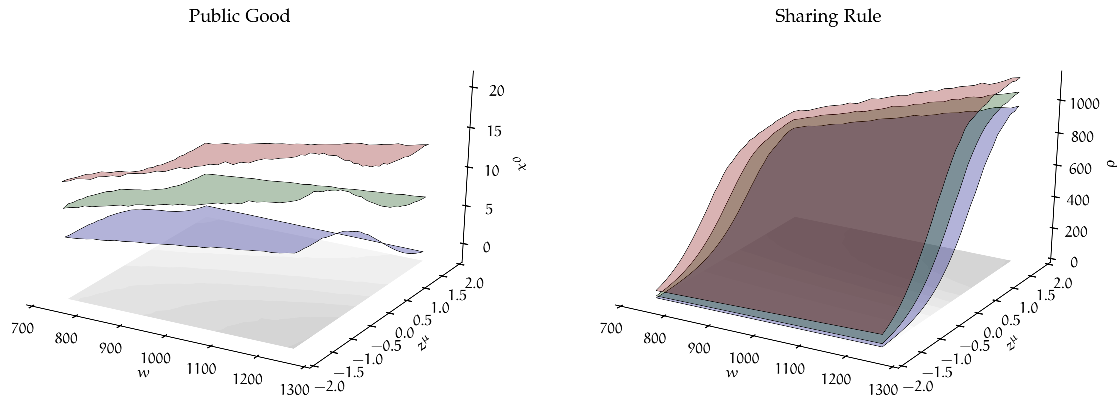

Figure 1.1: Data-Generating Process: Public

Figure 1.1 provides a limited view of the demand for the public good and the sharing rule; the components of the first stage demand system as a function of distribution factors z^μ and total endowment w. All other dimensions are set to a fixed value[9]. It is important to note that the function ρ in fact constitutes a local average over private errors $(\varepsilon_1^1, \varepsilon_2^1, \varepsilon_1^2, \varepsilon_2^2)$ as laid out in Section 1.4. As we have seen in equation (1.10), the sharing rule is a linear increasing function of total endowment w. Hence, the relative income allocation of the spouses ρ^h/ρ^w does not depend on total endowment, which is what we also see in real data and what is assumed in some identification strategies (see e.g. Dunbar et al. [2013b]). With respect to the distribution factor z^μ, we can see that the typical logistic shape is transformed into the sharing rule. An often used distribution factor in empirical work is the relative wage of the spouses. Based on this example, variation of the sharing rule with respect to this variable can be interpreted as follows. In case of two equal incomes, the endowment remaining after public consumption would be split equally among the spouses. As the incomes are shifted apart, this amount increases (decreases) for the spouse with higher (lower) income, although with decreasing magnitude as income becomes very unequal. Looking at the quantile curves, we can see that the way in which taste shocks for public goods enter demands generates heteroskedasticity in both public consumption and the sharing rule as a function of z^μ and w.

In what follows, we draw a sample of $N = 3{,}000$ households according to the data

[9] $(p_1^1, p_2^1, p_3^1, p_1^2, p_2^2, p_3^2) = (20, 20, 20, 20, 20, 20)$

generating process described above. We run 25 simulations and report the most important statistics (first and second moments, some quantiles and extrema, as well as bias and root mean squared errors) of the empirical distribution of the estimates and certain functions thereof. We do not estimate the parameters that determine the demands and sharing rule, but rather nonparametrically estimate the functions themselves. Hence we define a grid on the support of the exogenous variables and evaluate the estimated functions at each point, comparing them to their true values according to the data generating process. Functions are evaluated at an equally spaced grid of $G = 5$ points per dimension $K \in \{5, 9\}$. These estimates are aggregated to a scalar variable by means of the functions

$$\mathrm{MedRErr}(x) = Q^{0.5}_{j \in I_{G^K}} \left| \frac{\widehat{x}_j - x_j}{x_j} \right| \quad \text{and} \quad \mathrm{RASE}(x) = \left(\frac{1}{G^K} \sum_{j=1}^{G^K} (\widehat{x}_j - x_j)^2 \right)^{\frac{1}{2}}$$

defining median (absolute) relative error and the root average squared error. The unique minimum of these functions is zero and is attained if all estimated points are equal to the true function. In addition to this we calculate point-wise 95%-confidence bounds based on the true function value and evaluate how often the estimated demands fall within these bounds. Since the estimated demand system is a vector-valued function we count only those cases for which this is satisfied for all components of the estimate. We denote the confidence bounds for wife, husband and public as CI_w, CI_h and CI_0, respectively. For bandwith selection we use Silverman's rule of thumb making adjustments for the respective τ as proposed by Yu & Jones [1998]. All our simulations were executed on an EC2 Amazon Webservice `c3.8xlarge` instance with 36 virtual CPU's and 60 GB of memory.

In the first part of our simulation study we will focus on the estimated conditional quantiles of demand levels. To study both central tendency and tail behaviour of the function, we estimate its conditional median ($\tau = 0.5$) and 90$^{\text{th}}$ conditional quantile ($\tau = 0.9$).

Table 1.1 shows the simulation results for the conditional median using our proposed estimation procedure which takes private taste shocks into account when estimating public consumption and the sharing rule. Among the general properties of our estimator, we also want to investigate how it performs compared to a naive estimator, which omits these taste shocks. We have seen in the theoretical part that we would not be able to property identify the system without exploiting information about these shocks, using the residuals of private demands. Hence, we expect this naive estimator to be inconsistent in general.

	MedRErr(x_1^1)	MedRErr(x_2^1)	RASE(x_1^1)	RASE(x_2^1)	MedRErr(x_0)
True Values	-	-	-	-	-
Mean	0.0995	0.0672	1.7648	1.7252	0.0941
Median	0.0978	0.0675	1.7555	1.7030	0.0949
Std.Dev.	0.0066	0.0049	0.1050	0.1193	0.0067
10th Quantile	0.0924	0.0600	1.6351	1.5805	0.0854
90th Quantile	0.1094	0.0737	1.9096	1.8824	0.1043
Mean Bias	0.0995	0.0672	1.7648	1.7252	0.0941
Median Bias	0.0978	0.0675	1.7555	1.7030	0.0949
RMSE	0.0997	0.0674	1.7679	1.7293	0.0944
	MedRErr(ρ_h)	RASE(x_0)	RASE(ρ_h)	$\% \in CI_h$	$\% \in CI_0$
True Values	-	-	-	0.9500	0.9500
Mean	0.2159	1.0411	114.6111	0.9708	0.9251
Median	0.2154	1.0283	114.6837	0.9832	0.9320
Std.Dev.	0.0071	0.0482	2.0567	0.0269	0.0373
10th Quantile	0.2068	0.9862	111.8489	0.9335	0.8704
90th Quantile	0.2238	1.1076	116.7745	0.9972	0.9704
Mean Bias	0.2159	1.0411	114.6111	0.0208	-0.0249
Median Bias	0.2154	1.0283	114.6837	0.0332	-0.0180
RMSE	0.2160	1.0422	114.6296	0.0340	0.0449

Table 1.1: Private/Public Almost Ideal Demand Systems, $\tau = 0.5$, with integration

In contrast, Table 1.2 reports aggregate statistics of the naive estimator. In case of the conditional median, one can see that it actually performs equally well or even slightly better, if we look at the distance with respect to the true function (MedRErr, RASE). This stems from the fact that we have specified private taste shocks to be distributed according to a multivariate Normal with mean zero, which is symmetric, implying that the median with respect to each component is zero. By not including private taste shocks, we implicitly restrict them to be zero or in other words, assume that they are distributed according to a degenerate distribution with all mass at zero. Integrating with respect to this distribution creates less disturbance than integrating over the estimated (zero-mean) distribution for the residuals, while at the same time the expectation is not influenced due to (local) linearity.

	MedRErr(x_1^1)	MedRErr(x_2^1)	RASE(x_1^1)	RASE(x_2^1)	MedRErr(x_0)
True Values	-	-	-	-	-
Mean	0.0899	0.0594	1.5654	1.5313	0.0763
Median	0.0902	0.0590	1.5539	1.5175	0.0773
Std.Dev.	0.0097	0.0041	0.1558	0.1022	0.0143
10th Quantile	0.0768	0.0543	1.3195	1.4228	0.0591
90th Quantile	0.1032	0.0638	1.7775	1.6783	0.0966
Mean Bias	0.0899	0.0594	1.5654	1.5313	0.0763
Median Bias	0.0902	0.0590	1.5539	1.5175	0.0773
RMSE	0.0904	0.0596	1.5731	1.5347	0.0776
	MedRErr(ρ_h)	RASE(x_0)	RASE(ρ_h)	$\% \in CI_h$	$\% \in CI_0$
True Values	-	-	-	0.9500	0.9500
Mean	0.1828	0.8892	94.0917	0.9853	0.8586
Median	0.1854	0.9063	94.7524	0.9866	0.8600
Std.Dev.	0.0097	0.1397	3.4284	0.0111	0.0341
10th Quantile	0.1704	0.7158	89.3738	0.9682	0.8120
90th Quantile	0.1941	1.0929	97.6348	0.9985	0.8952
Mean Bias	0.1828	0.8892	94.0917	0.0353	-0.0914
Median Bias	0.1854	0.9063	94.7524	0.0366	-0.0900
RMSE	0.1831	0.9001	94.1542	0.0370	0.0976

Table 1.2: Private/Public Almost Ideal Demand Systems, $\tau = 0.5$, without integration

This result is not specific to our data generating process other than the symmetry assumption about the distribution of private taste shocks. Thus the naive estimator can consistently estimate the local expectation of the conditional median if it can be assumed that private taste shocks follow a symmetric distribution. The situation is generally different for the variance estimation, which is manifested in the underestimation of the width of the 95% confidence band. One special case, for which it would not make a difference would be the case where the slope with respect to taste shocks is one.

Next, we will consider the 90^{th} conditional quantile, to analyze how our estimation procedure works for estimating the tails of the distributions of taste and bargaining shocks.

	MedRErr(x_1^1)	MedRErr(x_2^1)	RASE(x_1^1)	RASE(x_2^1)	MedRErr(x_0)
True Values	-	-	-	-	-
Mean	0.0801	0.0984	2.0581	3.2167	0.1516
Median	0.0796	0.0954	2.0493	3.2122	0.1535
Std.Dev.	0.0074	0.0117	0.1803	0.2586	0.0128
10th Quantile	0.0714	0.0860	1.8249	2.8603	0.1372
90th Quantile	0.0879	0.1158	2.2595	3.5517	0.1681
Mean Bias	0.0801	0.0984	2.0581	3.2167	0.1516
Median Bias	0.0796	0.0954	2.0493	3.2122	0.1535
RMSE	0.0804	0.0991	2.0660	3.2271	0.1522
	MedRErr(ρ_h)	RASE(x_0)	RASE(ρ_h)	% $\in CI_h$	% $\in CI_0$
True Values	-	-	-	0.9500	0.9500
Mean	0.1019	2.1664	136.1037	0.9865	0.9702
Median	0.1007	2.1647	137.5085	0.9883	0.9760
Std.Dev.	0.0076	0.1436	5.8936	0.0102	0.0193
10th Quantile	0.0945	1.9874	129.6924	0.9724	0.9408
90th Quantile	0.1121	2.3478	143.0271	0.9981	0.9920
Mean Bias	0.1019	2.1664	136.1037	0.0365	0.0202
Median Bias	0.1007	2.1647	137.5085	0.0383	0.0260
RMSE	0.1022	2.1711	136.2312	0.0379	0.0279

Table 1.3: Private/Public Almost Ideal Demand Systems, $\tau = 0.9$, with integration

Table 1.3 again shows the correct estimator according to the theory. We see that this estimator performs quite well compared to what we have seen when estimating the conditional median. Naturally we see an increase in median absolute relative deviation and average squared errors, since estimating tails is more difficult due to the fact that the data becomes more sparse. While for the correct estimator the magnitude of this increase is within reasonable bounds, the picture changes drastically if we look at the results for the estimates of the 90^{th} conditional quantile of the naive estimator, which are reported in Table 1.4.

	MedRErr(x_1^1)	MedRErr(x_2^1)	RASE(x_1^1)	RASE(x_2^1)	MedRErr(x_0)
True Values	-	-	-	-	-
Mean	0.0716	0.0906	1.8744	2.9704	0.3144
Median	0.0721	0.0907	1.8924	3.0274	0.3117
Std.Dev.	0.0054	0.0106	0.1262	0.2374	0.0653
10th Quantile	0.0645	0.0778	1.6929	2.6782	0.2426
90th Quantile	0.0788	0.1025	2.0336	3.2637	0.4106
Mean Bias	0.0716	0.0906	1.8744	2.9704	0.3144
Median Bias	0.0721	0.0907	1.8924	3.0274	0.3117
RMSE	0.0718	0.0912	1.8787	2.9799	0.3211
	MedRErr(ρ_h)	RASE(x_0)	RASE(ρ_h)	$\% \in CI_h$	$\% \in CI_0$
True Values	-	-	-	0.9500	0.9500
Mean	0.1758	21.5328	245.4561	0.9934	1.0000
Median	0.1676	5.5092	142.6729	0.9958	1.0000
Std.Dev.	0.0481	77.0002	367.6604	0.0061	0.0000
10th Quantile	0.1320	4.2161	117.2577	0.9858	1.0000
90th Quantile	0.2130	7.6631	234.1738	0.9996	1.0000
Mean Bias	0.1758	21.5328	245.4561	0.0434	0.0500
Median Bias	0.1676	5.5092	142.6729	0.0458	0.0500
RMSE	0.1823	79.9543	442.0666	0.0439	0.0500

Table 1.4: Private/Public Almost Ideal Demand Systems, $\tau = 0.9$, without integration

To sum up, compared to what we found for the conditional median estimation, where not including private taste shocks did not effect consistency, omitting the latter when estimating the tails, induces a substantial bias. This is not surprising, since private taste shocks remain unexplained increasing the variance of the total unobserved variables and thus shifting the surfaces for conditional quantiles with $|\tau| > 0.5$ further apart. In addition to this, we see an overestimation of the conditional variance which is manifested in the too wide confidence bounds and also an immediate consequence of the inconsistent estimation.

Empirical Application

In order to illustrate our estimation procedure, we will now conduct an empirical study for which we use a sample of households from the Dutch LISS (Longitudinal Internet Studies for the Social Sciences) panel. This longitudinal study is collected by CentERdata and consists of 5000 households and 8000 individuals, which are drawn from the population register by Statistics Netherlands. Households that have no internet access are provided with the necessary hardware to participate in the study. We can thus assume that the sample is representative for the Dutch population. Our study is based on the *Time use and Consumption* module [Cherchye et al., 2015] which consists of two parts. In the time-use part of the study, respondents, consisting of all participants of the LISS core study that are at least 16 years old, were asked how much time they spent on a given category (e.g. leisure, labour supply, childcare-related activities) within the last seven days. The set of consumption-related questions concerns financial expenditures and is divided between what is defined as public, private and child-related consumption. In the first subset of consumption-related questions respondents had to allocate expenditures among 12 categories that can be argued to incur on a household level, including for example mortgage payments, rent, family trips but also formal childcare expenditures and public food consumption. The latter had to be assigned to the respective members (including children) which allows us to use it as part of private good consumption for our estimation. Private expenditure categories (9 in total) include for example expenditures for food and drinks that were consumed outside home and without family members, clothing and personal care. Child-related consumption expenditures (excluding food that was provided for and within the family) consists of the same categories as private goods for adults. The questions on child consumption are filled out by the parents if the child is less than 16 years old. For older children living at home, we take the amount of private consumption from the questionnaire of the respective child if the latter does not have his or her own funds arising from labour supply. If children have their own income and still live at home we do not treat their private consumption as part of the parents decision process, but rather treat it as an exogenous quantity for the parents and leave it to future research to consider children as endogenous decision-makers within this setting[10]. A detailed overview of the public and private consumption categories can be found in Appendix 1.C.

To complete our dataset we complement the study by data from the LISS core study, such as wages, age and education and pool all three available waves which were collected in 2010, 2011, 2013. This provides us with a sample of 743 observations, after dropping households in which at least one partner is not participating in the labour market as well as childless couples. Although we will sometimes refer to the partners as husband and wife, couples in our sample are not necessarily married, however they must live in a common household. We therefore define the population we are

[10]Dunbar et al. [2013a] estimate a collective model with children as decision makers.

interested in as the population of heterosexual couples with children, in which both partners participate in the labour market.

	mean	std	min	25%	50%	75%	max
Private Consumption	270.53	283.28	0.00	131.00	210.00	312.50	3100.00
Working Hours	25.37	11.61	1.00	18.00	24.00	32.00	96.00
Net Income	1116.93	592.33	0.00	750.00	1150.00	1423.50	4000.00
Age	43.16	7.68	22.00	37.00	44.00	49.00	62.00
Education	3.78	1.28	1.00	3.00	4.00	5.00	6.00
Food Consumption	131.41	102.83	0.00	85.00	125.00	165.00	2000.00
Private Consumption	237.44	305.86	0.00	105.00	175.00	280.00	4950.00
Working Hours	42.12	11.76	2.00	38.00	40.00	48.00	90.00
Net Income	2131.60	762.41	0.00	1703.50	2000.00	2450.00	10000.00
Age	45.69	7.89	26.00	40.00	46.00	52.00	87.00
Education	3.96	1.32	1.00	3.00	4.00	5.00	6.00
Food Consumption	130.76	105.26	0.00	82.50	120.00	160.00	2000.00
Public Consumption	1893.41	3008.40	0.00	1251.50	1732.50	2255.00	78796.00
Formal Childcare expenditure	97.12	254.37	0.00	0.00	0.00	33.75	2528.00
Public Food expenditure	428.43	302.76	0.00	300.00	400.00	550.00	6000.00
# of Children	2.02	0.80	1.00	2.00	2.00	2.00	6.00
Child Cons. (excl Food) ($\leqslant$16y)	137.68	203.38	0.00	0.00	90.00	181.00	2480.00
Child Cons. (excl Food) (>16y)	27.70	116.55	0.00	0.00	0.00	0.00	1720.00
Child Cons. (excl Food)	215.93	327.05	0.00	31.65	97.43	267.18	3687.02

Table 1.5: Wife, Husband and Household Descriptive Statistics (Raw)

Table 1.5 presents descriptive statistics of our dataset and is divided into three blocks for wife, husband and household level quantities, respectively. It can be seen that total private consumption (defined as the sum over all nine individual consumption sub-categories plus the assigned share of public food expenditure) is relatively low compared to what respondents report to be public consumption. The latter is also defined as an aggregate over all twelve categories, minus formal childcare expenditure and expenditure for food provided within the household. Looking at working hours, we can see that many women work part time, resulting in an average amount of 25.37 working hours (excluding commuting), compared to 42.12 hours for men. The picture is reversed for time spent on child care, where the women's share exceeds the men's with 13.88 hours compared to 8.86 hours. A substantial part of what we define as child consumption expenditure, namely child-related consumption (excluding food) plus public food consumption assigned to children, can be attributed to the latter.

In our collective labour supply model, we define composite private consumption and leisure as a private goods which is common in the literature [Chiappori, 1992; Fortin & Lacroix, 1997; Blundell et al., 2007; Cherchye et al., 2015]. Since we have information about total private consumption and time spent on leisure in our dataset, we can estimate private demands for all $s \in I_S = \{h, w\}$ by considering

$$[x^s, c^s](p_i^s, \rho_i^s, x_i^0, x_{c,i}^0, \varepsilon_i^s) = \arg\max_{x^s, c^s} \left\{ u^s(x^s, c^s, x_i^0, x_{c,i}^0, \varepsilon_i^s) \text{ s.t. } p_i^s x^s + c^s = \rho_i^s \;;\; 0 \leqslant x_i^s \leqslant T \right\}$$

as the individual consumption problem with leisure consumption x^s, corresponding price p^s which we define in terms of opportunity costs of labour supply (i.e. net hourly wage) and composite consumption good c^s whose price is normalized to one.

The public demands and the sharing rule $\left[x^0, x_c^0, \rho^w, \rho^h\right] \left(p_i^w, p_i^h, w_i, \varepsilon_i^h, \varepsilon_i^w, \varepsilon_i^0, \varepsilon_i^\mu\right)$ are the solution to the first stage allocation problem:

$$\begin{aligned}
\max_{x^0, x_c^0 0, \rho^w, \rho^h} \; & \nu^h(\rho^h, x^0, x_c^0, p_i^h, \varepsilon_i^h, \varepsilon_i^0)\mu(p_i^h, p_i^w, w_i, \varepsilon_i^\mu) \\
& \qquad + \nu^w(\rho^w, x^0, x_c^0, p_i^w, \varepsilon_i^w, \varepsilon_i^0)(1 - \mu(p_i^h, p_i^w, w_i, \varepsilon_i^\mu)) \\
\text{s.t. } & x^0 + x_c^0 + \rho^w + \rho^h = w_i
\end{aligned}$$

with composite public good x^0 and child-related good consumption which is treated as a public good x_c^0. We will assume that our structure (u^h, u^w, μ) satisfies Assumptions 1.3, 1.4 and 1.5 from Section 1.3 and Assumption 1.6 from Section 1.4. Unfortunately we do not have price variation for child-related consumption and public consumption. Hence we cannot identify substitution patterns with respect to price shocks other than the individual wages. We will however be able to identify income elasticities and the effects of distribution factors on both sharing and public good consumption.

In order to fit the structural model, we need to prepare our data by imposing some of the underlying restrictions. For each individual private demand system we impose the adding-up constraint by calculating individual endowment, represented by the sharing rule, as the expenditure for leisure plus consumption expenditure related to the private composite good. We define the maximum amount of labour by 126 hours per week, which corresponds to 18 hours per day, 7 days a week. Similarly, we also impose adding up for the first stage by calculating the household endowment w as the sum of public good consumption, child-related consumption and the shares ρ^w and ρ^h. We have two variables available which are often used as distribution factors, namely the age difference of the spouses and the wage ratio (p^w/p^h) which is assumed to not affect the budget contraint after controlling for total endowment.

	mean	std	min	25%	50%	75%	max
Leisure: x_1^w	86.75	16.80	26.00	77.50	89.00	98.00	123.00
Wage rate: p^w	11.25	10.62	0.00	7.70	10.16	12.91	190.53
Sharing Rule: ρ^w	1256.64	1064.42	0.00	853.28	1132.10	1455.65	17155.27
Leisure: x_1^h	75.02	14.07	5.00	68.00	76.00	84.00	122.00
Wage rate: p^h	13.07	11.45	0.00	9.35	11.26	14.30	259.82
Sharing Rule: ρ^h	1274.89	1342.58	129.39	843.84	1076.19	1378.24	30578.38
Child cons. x_0^c	215.93	327.05	0.00	31.65	97.43	267.18	3687.02
Wage ratio: p^w/p^h	1.10	2.09	0.00	0.60	0.90	1.20	50.00
Endowment: w	4716.69	3562.59	876.54	3548.27	4318.21	5206.00	80720.14

Table 1.6: Wife, Husband and Household Descriptive Statistics (Model)

Table 1.6 shows descriptive statistics of the constructed variables that are used for estimation. We can see that on average, after what is consumed publicly, remaining endowment is very equally distributed among the spouses, with a slightly higher mean for the husband which however mostly stems from very large observations. Looking at the quartiles, we can see that for the rest of the distribution the wife actually gets a slightly larger proportion of the household's endowment. The wage-ratio, which we use as distribution factor, is right-skewed with a median of 0.9 and a mean of 1.10, meaning that more than 50% (precisely 60%) of the women are out-earned by man, however we observe some households with very low male wage rates which are responsible for this long right tail.

Having defined the model variables, we specify our empirical model according to:

$$\begin{aligned} x_{1,i}^h &= x_1^h(p_i^h, \rho_i^h; \varepsilon_i^h) \\ x_{1,i}^w &= x_1^w(p_i^w, \rho_i^w; \varepsilon_i^w) \\ \left[x_c^0, \rho^w\right]_i &= \left[x_c^0, \rho^w\right]\left(p_i^w/p_i^h, w_i, \widehat{\varepsilon_i^h}, \widehat{\varepsilon_i^w}; \varepsilon_i^0, \varepsilon_i^\mu\right) \end{aligned}$$

By Theorem 1 the functions x_1^h, x_1^w and $\left[x_c^0, \rho^w\right]$ are monotone with respect to the unobserved taste and bargaining shocks ε_i^h, ε_i^w and $(\varepsilon_i^0, \varepsilon_i^\mu)$, respectively. Hence we can consistently estimate conditional quantiles and their derivatives using our proposed estimation procedure and calculate corresponding elasticities.

Note that for the public good and sharing rule, our specification involves the distribution factor p^w/p^h instead of p^w and p^h as separate variables. Since one is generally more interested in the effect of the distribution factor, rather than the wage-rates we impose this structure by making use of a standard homogeneity assumption on μ. This is not to say, that the unrestricted specification cannot be estimated. In fact, we estimated both models as a robustness check and it is comforting to report, that we found very little qualitative difference between the two specifications. One could even apply a formal test from which we will however abstract due to the fact that constructing such a test based on a bootstrap is computationally not feasible because of the high dimensionality. Deriving an analytic test statistic based on our asymptotic results would exceed the scope of this application, since to our knowledge a theory for uniform confidence bands for local polynomial quantile regression only exist for univariate models [Sabbah, 2014; Härdle & Song, 2009] such that one would have to develop a similar theory for a multivariate setting.

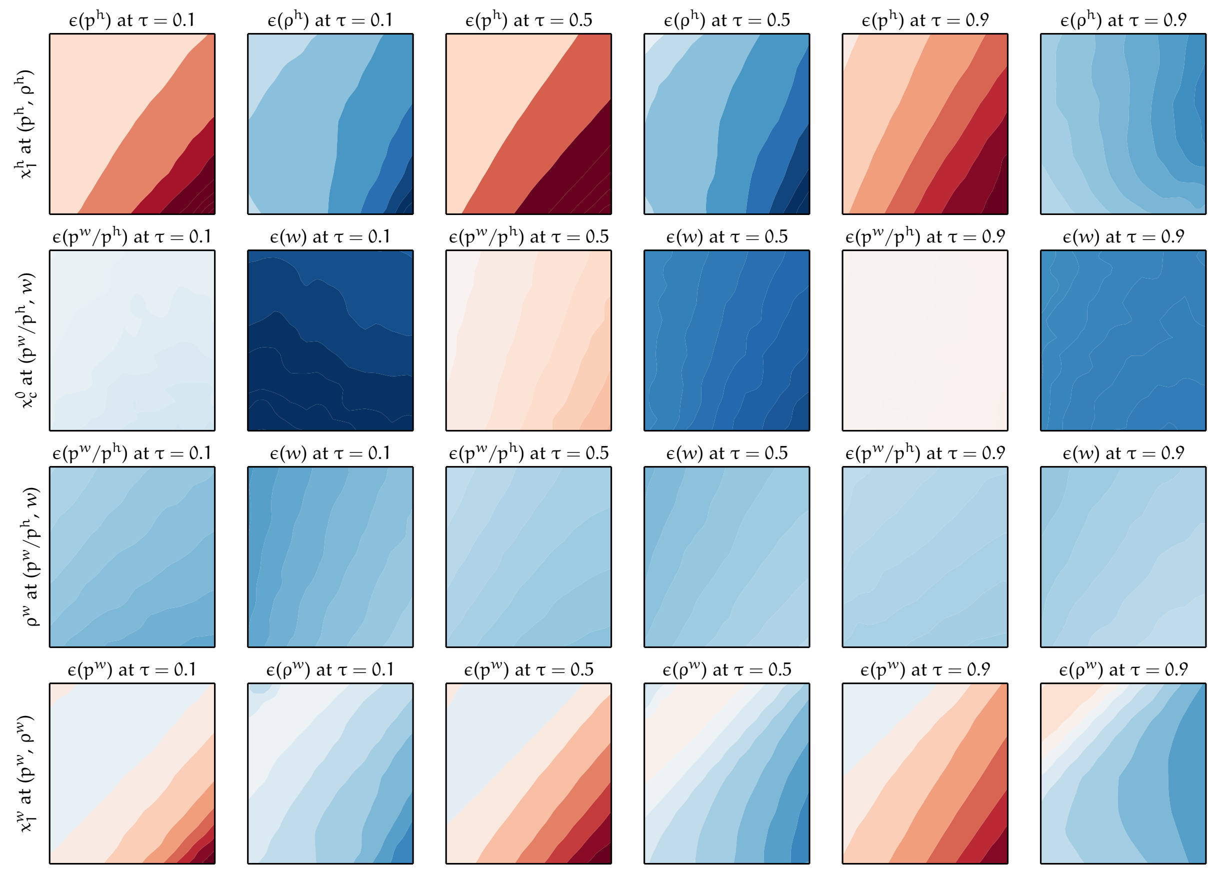

Figure 1.2: Surface plot: Elasticities ranging from -1.5 to $+1.5$ for $\tau \in \{0.1, 0.5, 0.9\}$

Figure 1.2 shows a heatmap of the estimated elasticities of both private and public demand systems. The surface color represents the estimate and ranges from dark orange to dark blue which are set to correspond to elasticities of -1.5 and $+1.5$, respectively. The major x-axis (3x2 columns) provides information about the exogenous variable whose variation is considered, as well as the considered quantile (τ). The major y-axis (4 rows) shows the endogenous variable the elasticity refers to, as well as the (minor) axis-labels for each graph in the grid. The interested reader can find tables with detailed estimates for elasticities with respect to all dimensions including standard errors for $\tau \in \{0.1, 0.25, 0.5, 0.75, 0.9\}$ evaluated at a grid representing the marginal quartiles of the exogenous variables in Appendix 1.B.

We will now briefly discuss some of the results. First, note that in general elasticities are not constant with respect to the exogenous variables, which can be seen from the color gradient of the respective graphs. While wage elasticities (columns 1,3 and 5) seem to be fairly linear in prices and endowment, where the latter is represented by the sharing rule, the influence of the sharing rule on private demands as well as the one of total endowment and the distribution factor on public demands, appears to be non-linear. We also note that elasticities are heterogeneous across the population, i.e. they depend on the realization with respect to the distribution of both taste and bargaining shocks which are indexed by τ. This heterogeneity mostly concerns the magnitude of the estimates, while the qualitative interpretation is the same across the population.

We start interpreting our results by looking at the wage elasticities for leisure x^h and x^w. It can be seen that for both spouses' ($s \in \{w, h\}$) demands they are negative, decreasing in p^s and increasing in ρ^s. While the negative sign of the elasticities is fairly obvious, the negative dependence on wage could be explained by the fact that the more an individual earns the more he or she can afford to adjust leisure (labour supply) to wage shocks. Increasing bargaining power on the other hand is associated with a smaller wage elasticity. One possible explanation for this is the associated higher share of total endowment as a secondary individual income source. Private endowment elasticities on the other hand are positive and only increasing with respect to wages, implying that the higher wages individuals receive the more they will increase their leisure when they face a positive endowment shock. Looking at different quantiles we can conclude that the magnitude of wage elasticities increases, the more taste one has for leisure, which seems intuitive.

Income elasticities on the other hand only increase for women, implying that the higher women valuate their leisure, the more rapid they adjust the latter to endowment shocks. Looking at the sharing rule, it is clear from the graph that the wife's share is significantly increasing with respect to the relative wage ratio. This finding is statistically and economically significant and in line with our expectations and with what is generally assumed in the literature. The corresponding elasticities seem to be constant with respect to both exogenous variables and bargaining shocks.

To disentangle husband's and wife's caring for children's utility, we consider the effect of the distribution factor (representing the wifes relative bargaining power) on child-related consumption. Although statistically not significantly different from zero, we find evidence which suggests that an increase in what we assume to have a positive effect on the wifes decision weight within the household actually decreases the amount of child-related expenditure. Since it is reasonable to assume that children's utilities increase in child-related consumption this would provide evidence against the hypothesis that is often made, stating that mothers care more for their children than fathers [Thomas, 1990; Lundberg et al., 1997]. While this result might seem surprising, it is consistent with what Cherchye et al. [2015] already found using the same dataset. It should be noted that we do not find this effect for the left tail of the distribution from which tastes for children are drawn. Finally we find that endowment elasticities for both child-related consumption and the sharing rule are positive and relatively constant with respect to relative wages and total endowment with values slightly decreasing in τ and ranging from 0.5 to 0.9, which is in line with what economic theory would suggest.

Since our identification strategy allows not only estimating conditional quantile of components of public and private demand systems, but also backing out unobserved taste preferences in this labour supply setting (see refinement in Section 1.3) in form of the distribution $\mathbf{P}$ characterizing the heterogeneous population, we can also answer questions about individual welfare. For example, for a given cardinalization of utilities, using a money-metric welfare measure we can estimate the distribution of a response to a policy measure of the entire population. In addition to this, given that we have conditional quantiles of sharing rules, one could construct Gini coefficients for men and women separately and evaluate how they would be effected by certain policies. Such a welfare analysis is however outside the scope of this paper and we will leave it for future research.

Conclusion

We have shown that if we have information on intra-household allocation, paired with the efficiency axiom in the collective household model that allows us to divide the within-group allocation process into two stages, we can not only identify the level of the sharing rule, a key ingredient of collective bargaining, but also introduce unobserved heterogeneity to the model. We gave necessary and sufficient conditions, which feature mostly separability restrictions between private and public goods in the underlying individuals' utility function, that allow us to exploit the two stage nature of the model for the purpose of dealing with heterogeneous agents. For this we proposed an estimation procedure, that allows to estimate a local average of the sharing rule and public consumption and proved its consistency and asymptotic normality. In the empirical part of the paper we conducted a Monte-Carlo analysis which provided evidence that considering private taste shocks is not only important from a theoretical point of view, but omitting them in the sharing rule and public good estimation leads to a substantial bias. We concluded the paper by estimating a collective labour supply model using the Dutch LISS panel and found that both functions and elasticities are highly non-linear in the exogenous variables. In addition to this, we came to the conclusion that quantile planes are not parallel which implies that the behaviour of individuals is heterogeneous across the population. Thus, it is clear that disregarding unobserved heterogeneity might lead to wrong conclusions.

One could think of a range of applications and extensions for the theory we developed in this paper. First, from an empirical point of view it would be very interesting to perform welfare analysis using conditional quantile estimates and the fact that we can recover the private taste preferences under the monotonicity or triangularity assumption, to predict how different parts of the population respond to policy measures in terms of both welfare and labour supply. Further, one could develop uniform confidence bounds for the estimated demands and derivatives to draw inference about whether or not preferences are of certain forms that are often imposed within the collective model (e.g. Browning et al. [2013]; Dunbar et al. [2013a]). One very special case would be the Gorman polar form [Gorman, 1953, 1961] with parallel Engle curves for each member, which would support the unitary model since individual preferences could then be aggregated to household or group preferences. In addition to this, the restrictions we derived might prove useful in approaches that estimate demands or recover the sharing rule implicitly such as Blundell et al. [2014] or Cherchye et al. [2015].

Auxiliary Results and Proofs

Lemma 1. *The key economic restrictions implied by the first order conditions* (1.6) *of the second stage problem can be expressed as*[11]

$$\Xi_i^s(x^s) := \frac{\nabla_{x^s_{-L_1}} u_i^s}{\nabla_{x^s_{L_1}} u_i^s} - \frac{p^s_{-L_1}}{p^s_{L_1}} = 0\,. \tag{1.7}$$

Similarly, the first stage restrictions (1.4)-(1.5) *implicitly defining the optimal allocation of public goods and the conditional sharing rule can be rewritten as the* $(L_0 + S - 1)$*-dimensional nonlinear system of equations*

$$\Xi_i^0(x_0, \rho) := \boldsymbol{\Omega}_i(x_0, \rho)\boldsymbol{\mu}_i - \mathbf{c} := \begin{bmatrix} \nabla_{x^0}\boldsymbol{\nu}_i^\mathsf{T} - \nabla_{\rho_1}\boldsymbol{\nu}_i^\mathsf{T} \otimes p^0 \\ \left(\frac{\partial \nu_{i,1}}{\partial \rho_1}\right)^{-1} J_0 \nabla_\rho \boldsymbol{\nu}_i \end{bmatrix} \boldsymbol{\mu}_i - \begin{bmatrix} 0_{L_0} \\ \iota_{S-1} \end{bmatrix} = 0\,. \tag{1.8}$$

where $\boldsymbol{\nu}_i = [\nu_i^1, \ldots, \nu_i^s]^\mathsf{T}$, $\boldsymbol{\rho}_i = [\rho_i^1, \ldots, \rho_i^s]^\mathsf{T}$, $\boldsymbol{\mu}_i = [1, \mu_i^2/\mu_i^1, \ldots, \mu_i^S/\mu_i^1]^\mathsf{T}$ *and* $(S-1) \times S$*-dimensional projection* $J_0 = \delta_{i,j+1}$ *with* δ *being defined as the Kronecker delta.*

Proof. The representation Ξ^s as defined in equation (1.7) is a standard result from individual utility maximization and this part of the proof is therefore left to the reader.

As for Ξ^0, let $\boldsymbol{\nu}_i = [\nu_i^1, \ldots, \nu_i^s]^\mathsf{T}$ and $\tilde{\boldsymbol{\mu}} = [\mu^1, \ldots, \mu^S]$. We write equation (1.5) in matrix notation omitting subscript $i \in I_N$ for simplicity

$$\nabla_{x^0}\boldsymbol{\nu}^\mathsf{T}\tilde{\boldsymbol{\mu}} - m^0 p^0 = 0_{L_0}$$

and substitute the Lagrange multiplier m^0 by $\mu^1 \frac{\partial \nu^1}{\partial \rho^1}$ without loss of generality using equation (1.4). Dividing both sides by μ^1, using the original definition $\boldsymbol{\mu} = [1, \mu^2/\mu^1, \ldots, \mu^S/\mu^1]^\mathsf{T}$, we get

$$\nabla_{x^0}\boldsymbol{\nu}^\mathsf{T}\boldsymbol{\mu} - \frac{\partial \nu^1}{\partial \rho^1} p^0 = 0_{L^0}.$$

As $\nabla_{\rho^s}\boldsymbol{\nu} = \left[\frac{\partial \nu^1}{\partial \rho^1}, 0, 0, \ldots\right]$ and the first element of $\boldsymbol{\mu}$ is 1, we can rewrite the second term to obtain

$$\nabla_{x^0}\boldsymbol{\nu}^\mathsf{T}\boldsymbol{\mu} - \left(\nabla_{\rho^1}\boldsymbol{\nu}^\mathsf{T} \otimes p^0\right)\boldsymbol{\mu} = 0_{L^0}.$$

Hence, the first part of the vector Ξ^0 representing L^0 rows follows. We have used the first of the S equations in (1.4). The remaining $S - 1$ equations can be written as

$$\mu^1 \frac{\partial \nu^1}{\partial \rho^1} = \mu^s \frac{\partial \nu^s}{\partial \rho^s}$$

for all $s = 2, \ldots, S$. Alternatively we can rewrite this as

$$\frac{\mu^s}{\mu^1}\left(\frac{\partial \nu^1}{\partial \rho^1}\right)^{-1} \frac{\partial \nu^s}{\partial \rho^s} - 1 = 0$$

[11]Subscript j refers to the j^{th} row of a vector, and $-j = \{j\}^c$ to all rows except j

and in matrix notation

$$\left[\left(\frac{\partial \nu^1}{\partial \rho^1}\right)^{-1} J_0 \nabla_\rho \boldsymbol{\nu}\right] \boldsymbol{\mu} - \iota_{S-1} = 0_{S-1}$$

with $(S-1) \times S$-dimensional projection $J_0 = \delta_{i,j+1}$ and δ being defined as the Kronecker delta. This projection removes the first row of the Jacobian matrix, taking into account that we only consider $s = 2, \ldots, S$ as we have only $S-1$ restrictions left. The term ι_{S-1} is a (constant) vector of ones, and as it does not influence the derivative of Ξ^0, which we will consider for our identification results, we will omit this. $\square$

Theorem 1 (Identification). *Let $\pi = [\pi^1, \ldots, \pi^S]$, $\Pi^0_\varepsilon = [\pi^0, \pi, \varepsilon]$ and $\tau \in (0,1)$. Under assumptions 1.1-1.5, both levels $Q^\tau_{x^s_j|\pi^s}(\pi^s)$ and derivatives $\dot{Q}^\tau_{x^s_j|\pi^s}(\pi^s)$ of the τ^{th} conditional quantile[12] of each component $j \in I_{L_1-1}$ of private demands are identified. In addition to this, for given individual private taste shocks ε, levels $Q^\tau_{(x^0_l,\rho)|\Pi^0_\varepsilon}(\Pi^0_\varepsilon)$ and derivatives $\dot{Q}^\tau_{(x^0,\rho)_l|\Pi^0_\varepsilon}(\Pi^0_\varepsilon)$ of each component $l \in I_{L_0+S-1}$ of public demands and the sharing rule are also identified.*

Proof. Define $\Pi^0_\varepsilon := (p^0, w, p, z, z^\mu, \varepsilon)$, $\Pi^s := (p^s, z^s)$ and $\varepsilon = [\varepsilon^1, \ldots, \varepsilon^S]$. Let us begin with the individual conditional optimization problem at stage two. As defined in (1.7), the first order conditions can be written as

$$\Xi^s(x^s) := \frac{\nabla_{x^s_{-L_1}} u^s}{\nabla_{x^s_{L_1}} u^s} - \frac{p^s_{-L_1}}{p^s_{L_1}}$$

for each $s \in I_S$ which does not depend on $\left[\varepsilon^0, \varepsilon^\mu\right]$ by Assumption 1.5. Using conventional properties of the individual's utility function (which have to hold for each ε^s), stated in Assumption 1.3.(i), for each $s \in I_S$ we can write

$$\nabla_{\varepsilon^s} x^s(\Pi^s, \varepsilon^s) = -\left[\nabla_{x^s} \Xi^s(x^s, \varepsilon^s)\right]^{-1} \nabla_{\varepsilon^s} \Xi^s(x^s, \varepsilon^s)$$

Monotonicity of each demand system $x^s(\Pi^s, \varepsilon^s)$ with respect to ε^s requires this Jacobian matrix to have full rank for all $s \in I_S$. Given that the inverse on the right hand side has full rank, a necessary and sufficient condition for this is that

$$\operatorname{rk}\left(\nabla_{\varepsilon^s} \Xi^s(x^s, \varepsilon^s)\right) = L_1 - 1$$

which follows from Assumption 1.3.(i). For a detailed treatment of this question in the unitary setting, which is equivalent to the second stage in the collective model, since we observe household sharing, see Beckert & Blundell [2008].

A similar reasoning can be applied to the first stage problem. For now, we treat ε as given and observed, and will later argue how to obtain an estimate for these individual taste shocks, using the first stage systems. As in equation (1.8) we can write the first order conditions as

$$\Xi^0\left(\left[x^0, \rho\right], \Pi^0_\varepsilon, \left[\varepsilon^0, \varepsilon^\mu\right]\right) = \Omega(x^0, \boldsymbol{\rho}, \Pi^0_\varepsilon, \varepsilon^0)\boldsymbol{\mu}(\Pi^0_\varepsilon, \varepsilon^\mu)$$

[12]The quantile of a r.v. Y conditional on X, with c.d.f. $F_{Y|X}$ is defined as $Q^\tau_y(x) := \inf\left\{y \in \mathbb{R} : F_{Y|X=x}(y) \geqslant \tau\right\}$.

with $(L_0 + S - 1) \times S$-matrix

$$\Omega(x^0, \boldsymbol{\rho}, \Pi^0_\varepsilon, \varepsilon^0) := \begin{bmatrix} \Omega_1 \\ \Omega_2 \end{bmatrix} = \begin{bmatrix} \nabla^T_{x^0} \boldsymbol{\nu}(x^0, \boldsymbol{\rho}, \Pi^0_\varepsilon, \varepsilon^0) - \nabla_{\rho_1} \boldsymbol{\nu}(x^0, \boldsymbol{\rho}, \Pi^0_\varepsilon, \varepsilon^0)^T \otimes p^0 \\ \left(\frac{\partial \nu_1(x^0, \rho_1, \Pi^0_\varepsilon, \varepsilon^0)}{\partial \rho_1}\right)^{-1} J_0 \nabla_\rho \boldsymbol{\nu}(x^0, \boldsymbol{\rho}, \Pi^0_\varepsilon, \varepsilon^0) \end{bmatrix}.$$

Once again, since by construction of the collective model $\nabla_{[x^0,\rho]}\Xi^0\left(\left[x^0, \rho\right], \Pi^0_\varepsilon, \left[\varepsilon^0, \varepsilon^\mu\right]\right)$ has full rank, we can write the observed system determining public demands and sharing rules for a given taste and bargaining-shock $\left[\varepsilon^0, \varepsilon^\mu\right]$ as the following:

$$\nabla_{[\varepsilon^0,\varepsilon^\mu]}\left[x^0, \rho\right]\left(\Pi^0_\varepsilon, \left[\varepsilon^0, \varepsilon^\mu\right]\right) = -\left[\nabla_{[x^0,\rho]}\Xi^0\left(\left[x^0, \rho\right], \Pi^0_\varepsilon, \left[\varepsilon^0, \varepsilon^\mu\right]\right)\right]^{-1} \nabla_{[\varepsilon^0,\varepsilon^\mu]}\Xi^0\left(\left[x^0, \rho\right], \Pi^0_\varepsilon, \left[\varepsilon^0, \varepsilon^\mu\right]\right).$$

In order to identify $\left[x^0, \rho\right]$ conditional upon ε, it remains to show that

$$\text{rk}\left(\nabla_{[\varepsilon^0,\varepsilon^\mu]}\Xi^0\left(\left[x^0, \rho\right], \Pi^0_\varepsilon, \left[\varepsilon^0, \varepsilon^\mu\right]\right)\right) = L_0 + S - 1$$

The derivative with respect to $\left[\varepsilon^0, \varepsilon^\mu\right]$ can be written as

$$\Psi := \nabla_{[\varepsilon^0,\varepsilon^\mu]}\Xi^0 = \left[\nabla_{\varepsilon^0}\Xi^0\ ,\ \nabla_{\varepsilon^\mu}\Xi^0\right] = \left[\left(\mu^T \otimes I_{L_0+S-1}\right)\nabla_{\varepsilon^0}\text{vec}\Omega\ ,\ \Omega\nabla_{\varepsilon^\mu}\boldsymbol{\mu}\right] \tag{1.11}$$

using the chain rule $\frac{d\text{vec}\Xi^0}{d\varepsilon^0} = \frac{d\text{vec}\Xi^0}{d\text{vec}\Omega}\frac{d\text{vec}\Omega}{d\varepsilon^0}$ in the second equality, noting that

$$\Xi^0 = \text{vec}\Xi^0 = \text{vec}\Omega\mu = \text{vec}I_{L_0+S-1}\Omega\mu = \left(\mu^T \otimes I_{L_0+S-1}\right)\text{vec}\Omega$$

and hence $\frac{d\Xi^0}{d\text{vec}\Omega} = \left(\mu^T \otimes I_{L_0+S-1}\right)$.

Now we can work out the first block of equation (1.11) and write it as the following $(L_0 + S - 1) \times L_0$-dimensional block matrix:

$$\left(\mu^T \otimes I_{L_0+S-1}\right)\nabla_{\varepsilon^0}\text{vec}\Omega := \begin{bmatrix} \Psi_{11} \\ \Psi_{21} \end{bmatrix} = \begin{bmatrix} \sum_{s\in I_S} \mu_s\left(\Pi^0_\varepsilon, \varepsilon^\mu_s\right)\nabla^2_{x^0,\varepsilon^0}\nu^s\left(x^0, \rho^s, \Pi^0_\varepsilon, \varepsilon^0\right) - \nabla^2_{\rho_1,\varepsilon^0}\boldsymbol{\nu}(x^0, \boldsymbol{\rho}, \Pi^0_\varepsilon, \varepsilon^0)^T \otimes p^0 \\ \boldsymbol{\mu}(\Pi^0_\varepsilon, \varepsilon^\mu)^T \nabla_{\boldsymbol{\varepsilon}^0}\left[\left(\frac{\partial \nu_1(x^0, \rho^1, \varepsilon^0)}{\partial \rho_1}\right)^{-1} J_0 \nabla_\rho \boldsymbol{\nu}(x^0, \boldsymbol{\rho}, \Pi^0_\varepsilon, \varepsilon^0)\right] \end{bmatrix}$$

with typical row f the second block for all $s = 2, \ldots, S$

$$(\cdot)_s = \frac{\mu_s(\Pi^0_\varepsilon, \varepsilon^0)}{\mu_1(\Pi^0_\varepsilon, \varepsilon^0)}\nabla_{\varepsilon^0}\left[\left(\frac{\partial \nu_1(x^0, \rho^1, \varepsilon^0)}{\partial \rho_1}\right)^{-1}\frac{\partial \nu_s(x^0, \rho^s, \varepsilon^0)}{\partial \rho_s}\right]$$

Using (1.11), we can now write Ψ as the block matrix

$$\Psi = \nabla_{[\varepsilon^0,\varepsilon^\mu]}\Xi^0 = \left[\nabla_{\varepsilon^0}\Xi^0\ ,\ \nabla_{\varepsilon^\mu}\Xi^0\right] = \begin{bmatrix} \Psi_{11} & \Psi_{12} \\ \Psi_{21} & \Psi_{22} \end{bmatrix}$$

with $\Psi_{j2} := \Omega_j\nabla_{\varepsilon^\mu}\boldsymbol{\mu}$. Since $\dim(\Psi_{11}) = L_0 \times L_0$ and $\dim(\Psi_{22}) = (S-1) \times (S-1)$, we can define the Schur complement

$$\Psi/\Psi_{22} = \Psi_{11} - \Psi_{12}\Psi_{22}^{-1}\Psi_{21}$$

and consider the following identity

$$\mathrm{rk}\,(\Psi) = \mathrm{rk}\,(\Psi_{22}) + \mathrm{rk}\,(\Psi/\Psi_{22})\,.$$

Regarding the rank of Ψ_{22}, we can again exploit the diagonal nature of the derivative $\nabla_{\varepsilon^\mu}\boldsymbol{\mu}$ and note that post-multiplying Ω with this diagonal matrix is equivalent to multiplying each of its rows componentwise with the same vector, namely, $\mathrm{diag}\,(\nabla_{\varepsilon^\mu}\boldsymbol{\mu})^T$ And since Ω has rank S and $\nabla_{\varepsilon^\mu}\boldsymbol{\mu}$ has rank $S-1$ it follows that $\mathrm{rk}\,(\Omega\nabla_{\varepsilon^\mu}\boldsymbol{\mu}) = S-1$, as long as $\frac{\partial\mu}{\partial\varepsilon_s^\mu} \neq 0$ for all $s \in I_S$, which follows from Assumption 1.3.(i). We can now use that for all $l \in I_{L_0}$ we have $\frac{\partial}{\partial\varepsilon_l}\left[\frac{\partial\nu^s}{\partial\rho^s}\Big/\frac{\partial\nu^{s'}}{\partial\rho^{s'}}\right] = 0$ for all $s, s' \in I_S$ by construction of μ and equation (1.4), which implies that $\Psi_{21} = 0$. In addition to this, we have $\frac{\partial^2\nu^{s'}}{\partial\rho^{s'}\varepsilon_l^0} = 0$ for some $s' \in I_S$ by Assumption 1.4, such that

$$\Psi_{11} = \sum_{s\in I_S}\mu_s\left(\Pi_\varepsilon^0, \varepsilon_s^\mu\right)\nabla^2_{x^0,\varepsilon^0}\nu^s\left(x^0, \rho^s, \Pi_\varepsilon^0, \varepsilon^0\right).$$

the Schur complement reduces to $\Psi/\Psi_{22} = \Psi_{11}$. To show that it has rank L_0, note that for each $s \in I_S$ the derivative $\nabla^2_{x^0,\varepsilon^0}\nu^s$ has full rank by Assumption 1.3.(i) and is diagonally dominant by Assumption 1.4. Thus, the derivative is positive semidefinite for each $s \in I_S$ and the linear combination as defined in Ψ_{11} has full rank as for any two (aligned) positive semidefinite matrices D_1, D_2 it holds that $\mathrm{rk}\,(D_1 + D_2) >= \min(\mathrm{rk}\,(D_1)\,, \mathrm{rk}\,(D_2))$. Note that the assumption $\frac{\partial^2\nu^{s'}}{\partial\rho^{s'}\varepsilon_l^0} = 0$ for some $s' \in I_S$ is a sufficient one and can be relaxed to $\nabla_{\rho^{s'},\varepsilon^0}\nu^{s'}$ not being a linear combination of the the columns of $\nabla_{x^0,\varepsilon^0}\nu^{s'}$, from which we could also conclude that $\mathrm{rk}\,(\Psi_{11}) = L_0$.

It follows that $\Psi = \nabla_{[\varepsilon^0,\varepsilon^\mu]}\Xi^0$ has full rank $L_0 + S - 1$ since we have shown that $\mathrm{rk}\,(\Psi) = \mathrm{rk}\,(\Psi_{11}) + \mathrm{rk}\,(\Psi_{22})$. Hence, also $\nabla_{[\varepsilon^0,\varepsilon^\mu]}\left[x^0, \rho\right]\left(\Pi_\varepsilon^0, \left[\varepsilon^0, \varepsilon^\mu\right]\right)$ has full rank which means that the resulting demand system $\left[x^0, \rho\right]$ is a monotonic function of $\left[\varepsilon^0, \varepsilon^\mu\right]$ and is thus identified. □

Lemma 5 (First stage Bahadur representation). *Let $K(\cdot)$ be a symmetric Kernel with bounded support and finite first derivative $\dot{K}(\cdot)$. Under Assumptions 1.1-1.6, as $h \to 0$ and $H_N = Nh^{L_1} \to \infty$:*

$$\sqrt{H_N}\begin{pmatrix}\widehat{Q^\tau_{x|\pi}}(\pi_0) - Q^\tau_{x|\pi}(\pi_0)\\ h(\widehat{\dot{Q}^\tau_{x|\pi}}(\pi_0) - \dot{Q}^\tau_{x|\pi}(\pi_0))\end{pmatrix} = D_0^{-1}(\pi_0)\frac{1}{\sqrt{Nh^K}}\sum_{i\in I_N}\left(\tau - \mathbb{1}\left\{x_i \leqslant \gamma^T z_i\right\}\right)\begin{pmatrix}1\\ \frac{\pi_i-\pi_0}{h}\end{pmatrix}K\left(\frac{\pi_i - \pi_0}{h}\right) + o_P(1)$$

Proof. Let $\gamma_0(\pi_0) = Q^\tau_{x|\pi}(\pi_0)$ and its derivative $\gamma_1(\pi_0) = \dot{Q}^\tau_{x|\pi}(\pi_0)$, with $\gamma = (\gamma_0, h\gamma_1)$. In addition to this, we define $z_i = \left(1, h^{-1}(\pi_i - \pi_0)^T\right)^T$ and $H_N = Nh^K$ where $K = L_1$ is the number of dimensions of π. Further let

$$g_n(\gamma) := \min_\gamma\sum_{i\in I_N}\rho_\tau\left(x_i - \gamma_0 - \gamma_1(\pi - \pi_0)\right)K\left(\frac{\pi_i - \pi_0}{h}\right) = \min_\gamma\sum_{i\in I_N}\left(\tau - \mathbb{1}\left\{x_i \leqslant \gamma^T z_i\right\}\right)z_i K\left(\frac{\pi_i - \pi_0}{h}\right)$$

and similarly using the law of iterative expectations its limit

$$g(\gamma) := \mathbb{E}\left(\tau - F_{x|\pi}(\gamma^T z)\right)zK\left(\frac{\pi - \pi_0}{h}\right) \tag{1.12}$$

with derivative

$$\dot{g}(\gamma) := -\mathbb{E}f_{x|\pi}(\gamma^\mathsf{T} z)\begin{pmatrix} 1 & h^{-1}(\pi-\pi_0)^\mathsf{T} \\ h^{-1}(\pi-\pi_0) & h^{-2}(\pi-\pi_0)(\pi-\pi_0)^\mathsf{T}\end{pmatrix} K\left(\frac{\pi-\pi_0}{h}\right),$$

which we use to define, slightly misusing notation denoting γ_0 as the true parameter,

$$D_0(\pi_0) := \dot{g}(\gamma_0) = -f_{x|\pi}(Q^\tau_{x|\pi}(\pi_0))f_\pi(\pi_0)\begin{pmatrix} \int K(u)du & \int u^\mathsf{T} K(u)du \\ \int uK(u)du & \int uu^\mathsf{T} K(u)du\end{pmatrix} = -f_{x|\pi}(Q^\tau_{x|\pi}(\pi_0))f_\pi(\pi_0)$$

by the change of variables $u = h^{-1}(\pi - \pi_0)$. The last equality follows from the normalization of the kernel to have zero mean and variance one. We know that γ_0 solves (1.12) and $\widehat{\gamma}_n$ solves (1.17). Thus,

$$\begin{aligned} 0 = g_n(\widehat{\gamma}_n) &= g_n(\widehat{\gamma}_n) - g(\widehat{\gamma}_n) - g_n(\gamma_0) + g(\gamma_0) + g(\widehat{\gamma}_n) + g_n(\gamma_0) \\ &= o_p(n^{-\frac{1}{2}}) + \dot{g}^\mathsf{T}(\gamma_0)(\widehat{\gamma}_n - \gamma_0) + (\dot{g}^\mathsf{T}(\xi) - \dot{g}^\mathsf{T}(\gamma_0))(\widehat{\gamma}_n - \gamma_0) + g_n(\gamma_0) \qquad (1.13) \\ &= o_p(n^{-\frac{1}{2}}) + D_0(\pi_0)(\widehat{\gamma}_n - \gamma_0) + o_p(\|\widehat{\gamma}_n - \gamma_0\|^2) + g_n(\gamma_0) \end{aligned}$$

where equation (1.13) follows from stochastic equicontinuity of the class containing g, noting that the kernel K has compact support and the class of directional derivatives ψ of the quantile loss function ρ_τ is a VC subgraph class with envelope function $2\max_{i\in I_N}|\pi_i|$, and thus Donsker [Kosorok, 2008; van der Vaart & Wellner, 1996]. In addition to this we apply the mean value theorem to $g(\widehat{\gamma}_n)$ around some $\xi \in [\gamma_0, \widehat{\gamma}_n]$ and use the Lipschitz assumption on $f_{x|\pi}$, from which we have $\|(\dot{g}^\mathsf{T}(\xi) - \dot{g}^\mathsf{T}(\gamma_0)\| \leqslant C\|\xi - \gamma_0\| \leqslant C\|\widehat{\gamma}_n - \gamma_0\|$. Hence we can write

$$\sqrt{H_N}(\widehat{\gamma}_n - \gamma_0) = -D_0^{-1}\sqrt{H_N}\, g_n(\gamma_0) + o_p(1)$$

from which, using the definition of g_n, the result follows. □

Lemma 2 (First stage asymptotic distribution)**.** *Let $K(\cdot)$ be a symmetric Kernel with bounded support and finite first derivative $\dot{K}(\cdot)$. Under Assumptions 1.1-1.6, as $h \to 0$ and $H_N := Nh^{L_1} \to \infty$:*

$$\sqrt{H_N}\begin{pmatrix} \widehat{Q^\tau_{x|\pi}}(\pi_0) - Q^\tau_{x|\pi}(\pi_0) \\ h(\widehat{\dot{Q}^\tau_{x|\pi}}(\pi_0) - \dot{Q}^\tau_{x|\pi}(\pi_0))\end{pmatrix} \rightsquigarrow \mathcal{N}\left(\frac{h^2}{2}B_0(\pi_0), \frac{\tau(1-\tau)B_1}{f_\pi(\pi_0)f^2_{x|\pi}(Q^\tau_{x|\pi}(\pi_0))}\right)$$

with

$$B_{0j}(\pi_0) = tr\left\{\ddot{Q}^\tau_{x|\pi}(\pi_0)\int uu^\mathsf{T}\begin{pmatrix}1\\u\end{pmatrix}_j K(u)\,du\right\}$$

and

$$B_1 = \begin{bmatrix} \int K^2(u)du & 0 \\ 0 & \int uu^\mathsf{T} K^2(u)du\end{bmatrix}.$$

Proof. We first consider the bias which is defined as:

$$
\begin{aligned}
&\mathbb{E}D_0(\pi_0)^{-1}\frac{1}{\sqrt{Nh^K}}\sum_{i\in I_N}\left(\tau-\mathbb{1}\left\{x_i\leqslant\gamma_0^T z_i\right\}\right)\begin{pmatrix}1\\ \frac{\pi_i-\pi_0}{h}\end{pmatrix}K\left(\frac{\pi_i-\pi_0}{h}\right)\\
&=D_0(\pi_0)^{-1}\frac{N}{\sqrt{Nh^K}}\mathbb{E}\left\{F_{x|\pi}(Q^\tau_{x|\pi}(\pi_i))-F_{x|\pi}\left(Q^\tau_{x|\pi}(\pi_0)+\dot{Q}^\tau_{x|\pi}(\pi_0)^T(\pi_i-\pi_0)\right)\right\}\begin{pmatrix}1\\ \frac{\pi_i-\pi_0}{h}\end{pmatrix}K\left(\frac{\pi_i-\pi_0}{h}\right)\\
&=-D_0(\pi_0)^{-1}\frac{\sqrt{Nh^K}}{h^K}\mathbb{E}\left\{\frac{h^2}{2}f_{x|\pi}(Q^\tau_{x|\pi}(\pi_0))\left(\frac{\pi_i-\pi_0}{h}\right)^T\ddot{Q}^\tau_{x|\pi}(\pi_0)\left(\frac{\pi_i-\pi_0}{h}\right)\right\}\begin{pmatrix}1\\ \frac{\pi_i-\pi_0}{h}\end{pmatrix}K\left(\frac{\pi_i-\pi_0}{h}\right)\\
&=-D_0(\pi_0)^{-1}\sqrt{Nh^K}\frac{h^2}{2}f_{x|\pi}(Q^\tau_{x|\pi}(\pi_0))f_\pi(\pi_0)\int u^T\ddot{Q}^\tau_{x|\pi}(\pi_0)u\begin{pmatrix}1\\ u\end{pmatrix}K(u)\,du\\
&=\sqrt{Nh^K}\frac{h^2}{2}B_0(\pi_0).
\end{aligned}
$$

We used the law of iterative expectations and the fact that $\tau=F_{x|\pi}(Q^\tau_{x|\pi}(\pi_i))$ in the first step. The second step follows from the mean value theorem of both $F_{x|\pi}$ terms around $Q^\tau_{x|\pi}(\pi_0)$, uniform boundedness and Lipschitz continuity of the density and the existence of a second derivative $\ddot{Q}^\tau_{x|\pi}(\pi_0)$:

$$
\begin{aligned}
&F_{x|\pi}(Q^\tau_{x|\pi}(\pi_i))-F_{x|\pi}\left(Q^\tau_{x|\pi}(\pi_0)+\dot{Q}^\tau_{x|\pi}(\pi_0)^T(\pi_i-\pi_0)\right)\\
&\quad=-f_{x|\pi}(Q^\tau_{x|\pi}(\pi_0))\left(Q^\tau_{x|\pi}(\pi_i)-Q^\tau_{x|\pi}(\pi_0)\right)-f_{x|\pi}(Q^\tau_{x|\pi}(\pi_0))\dot{Q}^\tau_{x|\pi}(\pi_0)^T(\pi_i-\pi_0))\\
&\quad=-\frac{1}{2}f_{x|\pi}(Q^\tau_{x|\pi}(\pi_0))(\pi_i-\pi_0)^T\ddot{Q}^\tau_{x|\pi}(\pi_0)(\pi_i-\pi_0).
\end{aligned}
$$

For the last step, we again change variables $u=h^{-1}(\pi_i-\pi_0)$ and use the fact that $u^TAu=\mathrm{tr}(Auu^T)$ for any square matrix A.

The variance of the element $g_n(\gamma_0)$ is defined as:

$$
\begin{aligned}
&\mathbb{E}D_0(\pi_0)^{-2}\frac{1}{Nh^K}\sum_{i\in I_N}\left(\tau-\mathbb{1}\left\{x_i\leqslant\gamma_0^T z_i\right\}\right)^2\begin{pmatrix}1 & h^{-1}(\pi_i-\pi_0)^T\\ h^{-1}(\pi_i-\pi_0) & h^{-2}(\pi_i-\pi_0)(\pi_i-\pi_0)^T\end{pmatrix}K\left(\frac{\pi_i-\pi_0}{h}\right)^2\\
&=\frac{\tau(1-\tau)}{h^K}D_0(\pi_0)^{-2}\mathbb{E}\begin{pmatrix}1 & h^{-1}(\pi_i-\pi_0)^T\\ h^{-1}(\pi_i-\pi_0) & h^{-2}(\pi_i-\pi_0)(\pi_i-\pi_0)^T\end{pmatrix}K\left(\frac{\pi_i-\pi_0}{h}\right)^2\\
&=\tau(1-\tau)\left(f_{x|\pi}(Q^\tau_{x|\pi}(\pi_0))\right)^{-2}(f_\pi(\pi_0))^{-1}\int\begin{pmatrix}1 & u^T\\ u & uu^T\end{pmatrix}K^2(u)\,du\\
&=\frac{\tau(1-\tau)B_1}{f^2_{x|\pi}(Q^\tau_{x|\pi}(\pi_0))f_\pi(\pi_0)}
\end{aligned}
$$

from which the asymptotic distribution follows together with Lemma 5. □

For the second stage we define $H_{0,N}=Nh^{K_0}$ and the functions

$$g_n(\theta,\gamma):=\sum_{i\in I_N}\left(\tau-\mathbb{1}\left\{\rho_i\leqslant\theta^T z_i(\gamma)\right\}\right)z_i(\gamma)K\left(\frac{\pi_i^0(\gamma)-\pi_0^0}{h}\right),$$

$$g(\theta,\gamma):=\mathbb{E}\left(\tau-F_{\rho|\pi^0}\left(\theta^T z_i(\gamma)\right)\right)z_i(\gamma)K\left(\frac{\pi_i^0(\gamma)-\pi_0^0}{h}\right)$$

with derivatives

$$\Gamma_\theta(\theta,\gamma) := \nabla_\theta g(\theta,\gamma) = -\mathbb{E}\left(f_{\rho|\pi^0}\left(\theta^T z_i(\gamma)\right)\right) z_i(\gamma) z_i^T(\gamma) K\left(\frac{\pi_i^0(\gamma)-\pi_0^0}{h}\right),$$

$$\begin{aligned}\Gamma_\gamma(\theta,\gamma) := \nabla_\gamma g(\theta,\gamma) = &-\mathbb{E}\left(f_{\rho|\pi^0}\left(\theta^T z_i(\gamma)\right)\right) z_i(\gamma)\theta^T K\left(\frac{\pi_i^0(\gamma)-\pi_0^0}{h}\right)\\ &+\mathbb{E}\left(\tau - F_{\rho|\pi^0}\left(\theta^T z_i(\gamma)\right)\right)\left\{K\left(\frac{\pi_i^0(\gamma)-\pi_0^0}{h}\right) I_K + z_i(\gamma)\mathrm{diag}\left\{\dot{K}\left(\frac{\pi_i^0(\gamma)-\pi_0^0}{h}\right)\right\}\right\}\end{aligned}$$

and constants $\Gamma_{\gamma,0} = \Gamma_\gamma(\theta_0,\gamma_0)$ and $\Gamma_{\theta,0} = \Gamma_\theta(\theta_0,\gamma_0)$.

Lemma 6 (Stochastic equicontinuity). *Under Assumptions 1.1-1.6, as* $h_0 \to 0$ *and* $H_{0,N} = N h_0^{L_0+SL_1+1} \to \infty$*:*

$$\sup_{(\theta,\gamma)\in\mathfrak{U}_\delta(\theta_0,\gamma_0)} \sqrt{H_{0,N}}\,\|g(\theta,\gamma) - g(\theta_0,\gamma_0) - g_n(\theta,\gamma) + g_n(\theta_0,\gamma_0)\| = o_p(1) \tag{1.14}$$

Proof. The function $g_n(\theta,\gamma)$ is a combination of Type II and Type IV in the sense of Andrews [1994]. In other words, we are faced with a non-differentiable objection function due to the quantile regression problem and in addition to this, our objective function for the second stage depends on the function we estimated in the first stage. Stochastic equicontinuity for such function is a well established result in the literature. For the i.i.d. case, *Theorem 3* in Chen et al. [2003] establishes stochastic equicontinuity for a class of functions with bracketing number satisfying $\int_0^\infty \sqrt{\log N(\epsilon^{1/s_j}, \mathcal{H}, \|\cdot\|_{\mathcal{H}})}d\epsilon < \infty$ for $j = 1,\dots,L_0 + SL_1 + 1$ and where $\mathcal{H}$ is a vector space of function endowed with pseudo-metric $\|\cdot\|_{\mathcal{H}}$. Our demand functions γ are members of this class of functions due to the smoothness assumption and van der Vaart & Wellner [1996, p. 154]. Let $m_{i,j}^\tau(\theta,\gamma) := \left(\tau - \mathbb{1}\left\{\rho_i \leqslant \theta^T z_i(\gamma)\right\}\right)\bar{z}_{i,j}(\gamma)$ where $\bar{z}_{i,j}(\gamma) = z_{i,j}(\gamma)K_i((\pi_i(\gamma) - \pi_0)/h)$. To check the conditions of the theorem we have to show that for some finite S_j and $s_j \in (0,1]$

$$\mathbb{E}\left(\sup_{(\theta,\gamma)\in\mathfrak{U}_\delta}\left|m_{i,j}^\tau(\theta,\gamma) - m_{i,j}^\tau(\theta_0,\gamma_0)\right|^r\right)^{\frac{1}{r}} \leqslant S_j\delta^{s_j},$$

where $\mathfrak{U}_\delta := \{(\theta,\gamma)\in\Theta\times\Gamma : \|\theta-\theta_0\|\leqslant\delta, \|\gamma-\gamma_0\|_{\mathcal{H}}\leqslant\delta\}$.

Hence, for any $(\theta,\gamma)\in\mathfrak{U}_\delta$ we can write the part inside the supremum as

$$\begin{aligned}&\left|m_{i,j}^\tau(\theta,\gamma) - m_{i,j}^\tau(\theta_0,\gamma_0)\right|\\ &\quad\leqslant \tau|\bar{z}_{i,j}(\gamma)-\bar{z}_{i,j}(\gamma_0)| + \left|\mathbb{1}\left\{\rho_i\leqslant\theta^T z_i(\gamma)\right\}\bar{z}_{i,j}(\gamma) - \mathbb{1}\left\{\rho_i\leqslant\theta_0^T z_i(\gamma_0)\right\}\bar{z}_{i,j}(\gamma_0)\right|\\ &\quad\leqslant \tau|\bar{z}_{i,j}(\gamma)-\bar{z}_{i,j}(\gamma_0)| + \left|\mathbb{1}\left\{\rho_i\leqslant\theta^T z_i(\gamma)\right\}\bar{z}_{i,j}(\gamma) - \mathbb{1}\left\{\rho_i\leqslant\theta^T z_i(\gamma)\right\}\bar{z}_{i,j}(\gamma_0)\right|\\ &\qquad + \left|\mathbb{1}\left\{\rho_i\leqslant\theta^T z_i(\gamma)\right\}\bar{z}_{i,j}(\gamma_0) - \mathbb{1}\left\{\rho_i\leqslant\theta_0^T z_i(\gamma_0)\right\}\bar{z}_{i,j}(\gamma_0)\right|\\ &\quad\leqslant \tau|\bar{z}_{i,j}(\gamma)-\bar{z}_{i,j}(\gamma_0)| + \left|\mathbb{1}\left\{\rho_i\leqslant\theta^T z_i(\gamma)\right\}\right||\bar{z}_{i,j}(\gamma)-\bar{z}_{i,j}(\gamma_0)|\\ &\qquad + \left|\mathbb{1}\left\{\rho_i\leqslant\theta^T z_i(\gamma)\right\} - \mathbb{1}\left\{\rho_i\leqslant\theta_0^T z_i(\gamma_0)\right\}\right||\bar{z}_{i,j}(\gamma_0)|\\ &\quad\leqslant 2|\bar{z}_{i,j}(\gamma)-\bar{z}_{i,j}(\gamma_0)| + \left|\mathbb{1}\left\{\rho_i\leqslant\theta^T z_i(\gamma)\right\} - \mathbb{1}\left\{\rho_i\leqslant\theta_0^T z_i(\gamma_0)\right\}\right||\bar{z}_{i,j}(\gamma_0)|.\end{aligned}$$

For $r \geqslant 2$ using monotonicity of $\bar{\rho} \mapsto \mathbb{1}\{\rho_i \leqslant \bar{\rho}\}$ we get for $\|\gamma - \gamma_0\|_{\mathcal{H}} \leqslant \delta$,

$$\begin{aligned}
&\mathbb{E}\left|m^{\tau}_{i,j}(\theta,\gamma) - m^{\tau}_{i,j}(\theta_0,\gamma_0)\right|^r \\
&\leqslant 2^r \mathbb{E}\left|\bar{z}_{i,j}(\gamma) - \bar{z}_{i,j}(\gamma_0)\right|^r + \mathbb{E}\left(\mathbb{1}\left\{\rho_i \leqslant \theta^{\top} z_i(\gamma)\right\} - \mathbb{1}\left\{\rho_i \leqslant \theta_0^{\top} z_i(\gamma_0)\right\}\right)\left|\bar{z}_{i,j}(\gamma_0)\right|^r \\
&\leqslant 2^r \mathbb{E} C_j^r \left|\gamma - \gamma_0\right|^r + \mathbb{E}\left(F_{\rho|\pi^0}(\theta^{\top} z_i(\gamma)) - F_{\rho|\pi^0}(\theta_0^{\top} z_i(\gamma_0))\right)\mathbb{E}\left|\bar{z}_{i,j}(\gamma)\right|^r \\
&\leqslant S_{j,1}\delta + S_{j,2}\delta
\end{aligned}$$

where the second last equation follows from the law of iterative expectations and the last equation from the fact that $(\gamma,\theta) \in \mathfrak{U}_\delta$, the moment conditions on z_i (Assumption 1.6) and the kernel $K(\cdot)$ and the smoothness assumption on $F_{\rho|\pi^0}$ which allows us to do a Taylor series expansion around (θ_0,γ_0) and use the fact that the density is uniformly bounded by some constant M. The local linear nature of our quantile regression function, and the construction of the "tick function" $m_{i,j}$ immediately imply Hölder continuity, the "natural" bounds $[0, x^{max}]$ for demands (for some finite constant x^{max}, implied by strictly positive prices and finite endowment) and the finiteness of the corresponding derivatives by Assumption 1.3.(i) imply compactness of the parameter space. Hence, all conditions in Chen et al. [2003] are satisfied, and stochastic equicontinuity of the process $\sqrt{H_{0,N}}(g(\cdot,\cdot) - g_n(\cdot,\cdot))$ as defined in equation (1.14) follows. □

Lemma 7 (Second stage consistency). *Under Assumptions 1.1-1.6, as* $h_0 \to 0$ *and* $H_{0,N} = N h_0^{L_0+SL_1+1} \to \infty$,

$$\left\|\widehat{\theta}_n - \theta_0\right\| = \mathcal{O}_p\left(H_{0,N}^{-\frac{1}{2}}\right)$$

Proof. Note that $g(\theta,\gamma)$ is differentiable for any θ and γ. Thus, a first order Taylor series expansion of around $\widehat{\theta}_n$ can be applied

$$g\left(\widehat{\theta}_n,\gamma_0\right) - g(\theta_0,\gamma_0) = \Gamma_{\theta,0}\left(\widehat{\theta}_n - \theta_0\right).$$

Taking the L^2-norm, a bound for the right hand side is obtained:

$$\left\|g\left(\widehat{\theta}_n,\gamma_0\right) - g(\theta_0,\gamma_0)\right\| \geqslant \lambda_{\min}(\Gamma_{\theta,0})\left\|\left(\widehat{\theta}_n - \theta_0\right)\right\|,$$

with $\lambda_{\min}(\Gamma_{\theta,0})$ being the the smallest eigenvalue of $\Gamma_{\theta,0}$. Since $g(\theta_0,\gamma_0) = 0$, it is sufficient to show that $\left\|g\left(\widehat{\theta}_n,\gamma_0\right)\right\| = \mathcal{O}_p\left(H_{0,N}^{-\frac{1}{2}}\right)$. Using the triangle inequality it follows that

$$\begin{aligned}
\left\|g\left(\widehat{\theta}_n,\gamma_0\right)\right\| &\leqslant \left\|g\left(\widehat{\theta}_n,\gamma_0\right) - g\left(\widehat{\theta}_n,\widehat{\gamma}_n\right)\right\| + \left\|g\left(\widehat{\theta}_n,\widehat{\gamma}_n\right)\right\| \\
&\leqslant \left\|g\left(\widehat{\theta}_n,\gamma_0\right) - g\left(\widehat{\theta}_n,\widehat{\gamma}_n\right)\right\| && (1.15)\\
&+ \left\|g\left(\widehat{\theta}_n,\widehat{\gamma}_n\right) - g(\theta_0,\gamma_0) - g_n\left(\widehat{\theta}_n,\widehat{\gamma}_n\right) + g_n(\theta_0,\gamma_0)\right\| && (1.16)\\
&+ \left\|g_n\left(\widehat{\theta}_n,\widehat{\gamma}_n\right)\right\| && (1.17)\\
&+ \left\|g_n(\theta_0,\gamma_0)\right\|, && (1.18)
\end{aligned}$$

where $g(\theta_0,\gamma_0) = 0$ was subtracted within the second norm (1.16). It is well established that the expression in equation (1.18) is tight: $\|g_n(\theta_0,\gamma_0)\| = \mathcal{O}_p\left(H_{0,N}^{-\frac{1}{2}}\right)$. The remaining equations (1.15), (1.16) and

(1.17) can again be analyzed separately. Starting with the first term, again using the triangle inequality, and changing the signs within the norm, equation (1.15) can be bounded by

$$\left\|g\left(\widehat{\theta}_n,\gamma_0\right)-g\left(\widehat{\theta}_n,\widehat{\gamma}_n\right)\right\| \leqslant \left\|g\left(\widehat{\theta}_n,\widehat{\gamma}_n\right)-g\left(\widehat{\theta}_n,\gamma_0\right)-\Gamma_\gamma\left(\widehat{\theta}_n,\gamma_0\right)(\widehat{\gamma}_n-\gamma_0)\right\| \tag{1.19}$$

$$+\left\|\Gamma_\gamma\left(\widehat{\theta}_n,\gamma_0\right)(\widehat{\gamma}_n-\gamma_0)-\Gamma_\gamma(\theta_0,\gamma_0)(\widehat{\gamma}_n-\gamma_0)\right\| \tag{1.20}$$

$$+\left\|\Gamma_\gamma(\theta_0,\gamma_0)(\widehat{\gamma}_n-\gamma_0)\right\|$$

Applying the Taylor series expansion of $g\left(\widehat{\theta}_n,\widehat{\gamma}_n\right)$ around γ_0 in equation (1.19) and using Lipschitz continuity of $\Gamma_\gamma\left(\widehat{\theta}_n,\gamma_0\right)$ with respect to both parameters, this reduces to

$$\left\|g\left(\widehat{\theta}_n,\gamma_0\right)-g\left(\widehat{\theta}_n,\widehat{\gamma}_n\right)\right\| \leqslant \mathcal{O}_p\left(\|\widehat{\gamma}_n-\gamma_0\|^2\right)+\mathcal{O}_p\left(\|\widehat{\gamma}_n-\gamma_0\|\left\|\widehat{\theta}_n-\theta_0\right\|\right)+\|\Gamma_{\gamma,0}(\widehat{\gamma}_n-\gamma_0)\| \tag{1.21}$$

$$=\|\Gamma_{\gamma,0}(\widehat{\gamma}_n-\gamma_0)\|(1+o_p(1))=\mathcal{O}_p\left(H_{0,N}^{-\frac{1}{2}}\right)$$

where the last equality follows from first stage consistency and the fact that $H_{0,N} < H_N$. The term (1.16) is $o_p(H_{0,N}^{-1/2})$ by Lemma 6. Thus, we have

$$\lambda_{\min}(\Gamma_{\theta,0})\left\|\left(\widehat{\theta}_n-\theta_0\right)\right\| \leqslant \left\|g\left(\widehat{\theta}_n,\gamma_0\right)\right\|=\mathcal{O}_p\left(H_{0,N}^{-\frac{1}{2}}\right)$$

which completes the proof. □

Lemma 3 (Bahadur representation 2nd stage). *Let* $K(\cdot)$ *be a symmetric Kernel with bounded support and finite first derivative* $\dot{K}(\cdot)$*. Under Assumptions 1.1-1.6, as* $h, h_0 \to 0$ *and* $H_N, H_{0,N} \to \infty$*:*

$$\sqrt{H_{0,N}}\begin{pmatrix}\widehat{Q^\tau_{\rho|\pi^0}}(\pi^0_0)-Q^\tau_{\rho|\pi^0}(\pi^0_0)\\ h_0(\widehat{\dot{Q}^\tau_{\rho|\pi^0}}(\pi^0_0)-\dot{Q}^\tau_{\rho|\pi^0}(\pi^0_0))\end{pmatrix}=$$

$$-\Gamma_{\theta,0}^{-1}\frac{1}{\sqrt{H_{0,N}}}\sum_{i\in I_N}\left(\tau-\mathbb{1}\left\{\rho_i\leqslant\theta^T z_i^0(\gamma)\right\}\right)\begin{pmatrix}1\\ \frac{\pi_i^0-\pi_0^0}{h_0}\end{pmatrix}K\left(\frac{\pi_i^0-\pi_0^0}{h_0}\right)$$

$$+\sqrt{\frac{H_{0,N}}{H_N}}\Gamma_{\theta,0}^{-1}\Gamma_{\gamma,0}D_0^{-1}\frac{1}{\sqrt{H_N}}\sum_{i\in I_N}\left(\tau-\mathbb{1}\left\{x_i\leqslant\gamma^T z_i\right\}\right)\begin{pmatrix}1\\ \frac{\pi_i-\pi_0}{h}\end{pmatrix}K\left(\frac{\pi_i-\pi_0}{h}\right)+o_P(1)$$

Proof. We first show that we can rewrite

$$g_n\left(\widehat{\theta}_n,\widehat{\gamma}_n\right)=g_n(\theta_0,\gamma_0)+\Gamma_{\theta,0}(\widehat{\theta}_n-\theta_0)+\Gamma_{\gamma,0}(\widehat{\gamma}_n-\gamma_0)+o_p\left(H_{0,n}^{-\frac{1}{2}}\right).$$

Thus, by adding and subtracting we get

$$g_n\left(\widehat{\theta}_n,\widehat{\gamma}_n\right)-g_n(\theta_0,\gamma_0)-\Gamma_{\theta,0}(\widehat{\theta}_n-\theta_0)-\Gamma_{\gamma,0}(\widehat{\gamma}_n-\gamma_0)$$

$$=g_n\left(\widehat{\theta}_n,\widehat{\gamma}_n\right)-g_n(\theta_0,\gamma_0)-\Gamma_{\theta,0}(\widehat{\theta}_n-\theta_0)-\Gamma_{\gamma,0}(\widehat{\gamma}_n-\gamma_0)$$

$$+g\left(\widehat{\theta}_n,\widehat{\gamma}_n\right)-g\left(\widehat{\theta}_n,\widehat{\gamma}_n\right)+g(\theta_0,\gamma_0)-g(\theta_0,\gamma_0)$$

$$+g\left(\widehat{\theta}_n,\gamma_0\right)-g\left(\widehat{\theta}_n,\gamma_0\right)+\Gamma_\gamma\left(\widehat{\theta}_n,\gamma_0\right)(\widehat{\gamma}_n-\gamma_0)-\Gamma_\gamma\left(\widehat{\theta}_n,\gamma_0\right)(\widehat{\gamma}_n-\gamma_0)+o_p\left(H_{0,n}^{-\frac{1}{2}}\right).$$

Taking norms, rearranging the terms on the right hand side, using the triangle inequality, the following bound is obtained for:

$$\begin{aligned}
&\left\| g_n\left(\widehat{\theta}_n, \widehat{\gamma}_n\right) - \left(g_n\left(\theta_0, \gamma_0\right) + \Gamma_{\theta,0}(\widehat{\theta}_n - \theta_0) + \Gamma_{\gamma,0}\left(\widehat{\gamma}_n - \gamma_0\right)\right)\right\| \\
&\leqslant \left\| g_n\left(\widehat{\theta}_n, \widehat{\gamma}_n\right) - g_n\left(\theta_0, \gamma_0\right) - \left(g\left(\widehat{\theta}_n, \widehat{\gamma}_n\right) - g\left(\theta_0, \gamma_0\right)\right)\right\| \\
&+ \left\| g\left(\widehat{\theta}_n, \widehat{\gamma}_n\right) - g\left(\widehat{\theta}_n, \gamma_0\right) - \Gamma_\gamma\left(\widehat{\theta}_n, \gamma_0\right)\left(\widehat{\gamma}_n - \gamma_0\right)\right\| \\
&+ \left\| g\left(\widehat{\theta}_n, \gamma_0\right) - g\left(\theta_0, \gamma_0\right) - \Gamma_{\theta,0}\left(\widehat{\theta}_n - \theta_0\right)\right\| \\
&+ \left\| \Gamma_\gamma\left(\widehat{\theta}_n, \gamma_0\right)\left(\widehat{\gamma}_n - \gamma_0\right) - \Gamma_{\gamma,0}\left(\widehat{\gamma}_n - \gamma_0\right)\right\| = o_p\left(H_{0,N}^{-\frac{1}{2}}\right),
\end{aligned}$$

using equations (1.19), (1.20), (1.21), (1.14), and $\sqrt{H_{0,N}}$-consistency of $\widehat{\theta}_n$.

The first order condition $g_n\left(\theta, \widehat{\gamma}_n\right) = 0$ is solved by $\widehat{\theta}_n$ such that

$$0 = g_n\left(\widehat{\theta}_n, \widehat{\gamma}_n\right) = g_n\left(\theta_0, \gamma_0\right) + \Gamma_{\theta,0}(\widehat{\theta}_n - \theta_0) + \Gamma_{\gamma,0}\left(\widehat{\gamma}_n - \gamma_0\right) + o_p\left(H_{0,N}^{-\frac{1}{2}}\right).$$

Since $\Gamma_{\theta,0}$ has full rank, by premultiplying $\sqrt{H_{0,N}}$, an asymptotic representation of the second stage estimator is obtained

$$\sqrt{H_{0,N}}\left(\widehat{\theta}_n - \theta_0\right) = -\sqrt{H_{0,N}}\Gamma_{\theta,0}^{-1} g_n\left(\theta_0, \gamma_0\right)) + \sqrt{H_{0,N}/H_N}\Gamma_{\theta,0}^{-1}\Gamma_{\gamma,0}\sqrt{H_N}\left(\widehat{\gamma}_n - \gamma_0\right) + o_p(1)$$

which, using the definition of g_n and the asymptotic distribution of the first stage gives us the Bahadur representation. □

Lemma 8 (Conditions for Thm 2.8.9 in van der Vaart & Wellner [1996]). *Let* $\rho_P(f, g) = \|(f - g) - P(f - g)\|_{P,2}$. *Under assumptions 1.1-1.6, we have that for all* $\theta_1, \theta_2 \in \Theta$, *where* Θ *satisfies the uniform entropy condition* $\widehat{\mathbf{P}}_n F^2 = O(1)$,

(i) $\lim_{n\to\infty} \sup_{\theta_1,\theta_2} \left| \rho_{\widehat{P}_n}(\theta_1, \theta_2) - \rho_{P_0}(\theta_1, \theta_2)\right| = 0$

(ii) $\lim_{n\to\infty} \mathbb{E}\left(\widehat{a}_{H_N}^2 \mathbb{1}\{\widehat{a}_{H_N} \geqslant \sqrt{n}\varepsilon\}\right) = 0\,.$

Proof. For the first part, let $\overline{\theta} = \mathrm{vec}\left([\theta_1, -\theta_2]\right)$ and $\theta_1 - \theta_2 = (I_K \otimes \iota_K^T)\overline{\theta}$ such that $\mathrm{Var}\left[\theta_1 - \theta_2\right] = (I_K \otimes \iota_K^T)\mathrm{Var}\left[\overline{\theta}\right](I_K \otimes \iota_K)$.

Then we can write the distance between θ_1 and θ_2 according to the semi-metric defined above as

$$
\begin{aligned}
&\left|\rho_{\widehat{P}_n}(\theta_1,\theta_2)-\rho_{P_0}(\theta_1,\theta_2)\right| = |\mathrm{Var}\,[\theta_1(\widehat{a}_{H_N})-\theta_2(\widehat{a}_{H_N})] - \mathrm{Var}\,[\theta_1(a)-\theta_2(a)]| \\
&= \left|(\iota_2^T\otimes I_K)\left\{\mathrm{Var}\left[\overline{\theta}(\widehat{a}_{H_N})\right]-\mathrm{Var}\left[\overline{\theta}(a)\right]\right\}(\iota_2\otimes I_K)\right| \\
&= \left|(\iota_2^T\otimes I_K)\left\{\mathrm{Var}\left[\overline{\theta}(a)+\dot{\overline{\theta}}(a)(\widehat{a}_{H_N}-a)\right]-\mathrm{Var}\left[\overline{\theta}(a)\right]\right\}(\iota_2\otimes I_K)\right| \\
&= \left|(\iota_2^T\otimes I_K)\left\{\mathrm{Var}\left[(\iota_2^T\otimes I_{2K})\mathrm{vec}\left([\overline{\theta}(\widehat{a}_{H_N}),\dot{\overline{\theta}}(a)(\widehat{a}_{H_N}-a)\right)\right]-\mathrm{Var}\left[\overline{\theta}(a)\right]\right\}(\iota_2\otimes I_K)\right| \\
&= \left|(\iota_2^T\otimes I_K)(\iota_2^T\otimes I_{2K})\left\{\begin{bmatrix}\mathrm{Var}\left[\overline{\theta}(a)\right] & \zeta(a,\widehat{a}_{H_N}) \\ \zeta(a,\widehat{a}_{H_N}) & \mathrm{Var}\left[\dot{\overline{\theta}}(a)(\widehat{a}_{H_N}-a)\right]\end{bmatrix}-\begin{bmatrix}\mathrm{Var}\left[\overline{\theta}(a)\right] & 0 \\ 0 & 0\end{bmatrix}\right\}(\iota_2\otimes I_{2K})(\iota_2\otimes I_K)\right| \\
&= \left|(\iota_2^T\otimes I_K)(\iota_2^T\otimes I_{2K})\left\{\begin{bmatrix}0 & \zeta(a,\widehat{a}_{H_N}) \\ \zeta(a,\widehat{a}_{H_N}) & \mathrm{Var}\left[\dot{\overline{\theta}}(a)(\widehat{a}_{H_N}-a)\right]\end{bmatrix}\right\}(\iota_2\otimes I_{2K})(\iota_2\otimes I_K)\right| = 0
\end{aligned}
$$

since both $\mathrm{Var}\left[\dot{\overline{\theta}}(a)(\widehat{a}_{H_N}-a)\right]=0$ and $\zeta(a,\widehat{a}_{H_N})=\mathrm{Cov}\left[\overline{\theta}(a),\dot{\overline{\theta}}(a)(\widehat{a}_{H_N}-a)\right]=0$ which follows from consistency of the first stage.

For the second condition, recall that $\widehat{a}_{H_N}=x-\widehat{Q^\tau_{x|\pi}}(\pi_0)$. Thus, for positive prices and finite budgets characterized by π_0 we get a natural bound of $|\widehat{a}_{H_N}|\leqslant x^{\max}<\infty$. The domain with respect to first stage estimated taste shocks of the quantile functions characterizing public demands θ is thus compact. Together with the fact that $|\nabla_{\varepsilon^s}x^s|\neq 0$ on that domain, smoothness of θ and finite moments of ε^s for $s\in I_S$, we get a finite envelope of the function space. Thus the entropy condition is satisfied. Since the envelope is a constant, the Lindeberg condition given by Lemma 8.(i) is also satisfied. $\square$

Theorem 2 (Asymptotic distribution of local average conditional quantile)**.** *Let* $a_i^*\sim\widehat{F}_{\widehat{a}}$ *for* $i\in I_n$ *and* $H_{0,N}=Nh_0^{L_0+SL_1+1}$ *where* $\mathcal{O}(H_N)=o(n)$. *Then under Assumptions 1.1-1.6, as* $h_0\to 0$ *and* $H_{0,N}\to\infty$,

$$
\sqrt{H_{0,N}}\begin{pmatrix} n^{-1}\sum_{i\in I_n}\widehat{Q^\tau_{\rho|\pi^0}}(\pi_0^0,a_i^*)-\int Q^\tau_{\rho|\pi^0}(\pi_0^0,a)dF_a(a) \\ n^{-1}\sum_{i\in I_n}h_0(\widehat{\dot{Q}^\tau_{\rho|\pi^0}}(\pi_0^0,a_i^*)-\int \dot{Q}^\tau_{\rho|\pi^0}(\pi_0^0,a)dF_a(a))\end{pmatrix}\rightsquigarrow\mathcal{N}\left(\mathbb{B}(\pi_0^0),\mathbb{V}(\pi_0^0)\right)
$$

with $\mathbb{B}(\pi_0^0)=\frac{h_0^2}{2}B_0^0\int\ddot{Q}^\tau_{\rho|\pi^0}\left(\pi_0^0(a)\right)dF_a(a)$ *and* $\mathbb{V}(\pi_0^0)$ *defined in equations* (1.24) *and* (1.25) *at the end the proof.*

Proof. Let $\widehat{\mathbb{P}}_n$ and $\widehat{\mathbf{P}}_n$ be the sample and the true probability measure of $\widehat{a}$ respectively. Similarly denote $\mathbb{P}_n$ and $\mathbf{P}_0$ the ones for the true underlying taste shock a. Further we define the laws $\mathbb{G}_{n,P_n}=\sqrt{n}(\widehat{\mathbb{P}}_n-\widehat{\mathbf{P}}_n)$ and $\mathbb{G}_{n,P_0}=\sqrt{n}(\mathbb{P}_n-\mathbf{P}_0)$, respectively. We start by expanding our object of interest as follows:

$$
\begin{aligned}
\sqrt{n}\left(\widehat{\mathbb{P}}_n\widehat{\theta}_n-\mathbf{P}_0\theta_0\right) &= \sqrt{n}\left(\widehat{\mathbb{P}}_n\widehat{\theta}_n-\widehat{\mathbf{P}}_n\widehat{\theta}_n\right)+\sqrt{n}\left(\widehat{\mathbf{P}}_n\widehat{\theta}_n-\mathbf{P}_0\widehat{\theta}_n\right)+\sqrt{n}\left(\mathbf{P}_0\widehat{\theta}_n-\mathbf{P}_0\theta_0\right) \\
&= \mathbb{G}_{n,P_n}\theta_0+\mathbb{G}_{n,P_n}\left(\widehat{\theta}_n-\theta_0\right)+\sqrt{n}\left(\mathbf{P}_n-\mathbf{P}_0\right)\widehat{\theta}_n+\mathbf{P}_0\sqrt{n}\left(\widehat{\theta}_n-\theta_0\right). \qquad (1.22)
\end{aligned}
$$

By Theorem 2.8.9 in van der Vaart & Wellner [1996] and Lemma 8 the distribution of $\mathbb{G}_{n,P_n}\theta_0$ in equation

(1.22) is the same as $\mathbb{G}_{n,P_0}\theta_0$ as $n \to \infty$. Thus, we have

$$\begin{aligned} \sqrt{n}\left(\widehat{\mathbb{P}}_n\widehat{\theta}_n - \mathbf{P}_0\theta_0\right) &= \mathbb{G}_{n,P_0}\theta_0 + \sqrt{n}\left(\mathbb{P}_n\widehat{\theta}_n - \mathbf{P}_0\widehat{\theta}_n - \mathbb{P}_n\theta_0 - \mathbf{P}\theta_0\right) \\ &\quad + \sqrt{n}\left(\widehat{\gamma}_n - \widehat{\gamma}\right)\left(\mathbf{P}_n - \mathbf{P}_0\right)C_a + \mathbf{P}_0\sqrt{n}\left(\widehat{\theta}_n - \theta_0\right) \\ &= \mathbb{G}_{n,P_0}\theta_0 + \mathbf{P}_0\sqrt{n}\left(\widehat{\theta}_n - \theta_0\right) + o_p(1) \end{aligned} \tag{1.23}$$

where the constant C_a is defined by the derivative of θ with respect to γ, which is the only thing that depends on a over which we take expectations. In the last equation we use the rate restriction on n and relate it to $H_{0,N}$ and consistency of the first stage. Since demand functions θ are smooth in a, or more precisely it holds that $2\alpha > d$ where α is the order of the differential operator and d is the dimension of a which is $d = S(L_1 - 1)$ in this case. Thus, the function space for both its true value θ_0 and its estimates $\widehat{\theta}_n$ is P_0-Donsker [van der Vaart & Wellner, 1996, Example 2.10.25, p.202] and as a result stochastic equicontinuity for the second term in equation (1.23) holds (see van der Vaart & Wellner [1996, Section 2.1.2]).

For asymptotic normality we use the linearization as defined above and pick n such that $o(n) = \mathcal{O}(H_N)$ which is satisfied by the slower rate of the second stage $H_{N,0}$:

$$\begin{aligned} \sqrt{n}\left(\widehat{\mathbb{P}}_n\widehat{\theta}_n - \mathbf{P}_0\theta_0\right) &= \mathbb{G}_{n,P_0}\theta_0 + \sqrt{n/H_{0,N}}\mathbf{P}_0\sqrt{H_{0,N}}\left(\widehat{\theta}_n - \theta_0\right) \\ &= \mathbb{G}_{n,P_0}\theta_0 + \mathbf{P}_0\left\{-\sqrt{H_{0,N}}\Gamma_{\theta,0}^{-1}g_n\left(\theta_0,\gamma_0\right)) + \sqrt{H_{0,N}/H_N}\Gamma_{\theta,0}^{-1}\Gamma_{\gamma,0}\sqrt{H_N}\left(\widehat{\gamma}_n - \gamma_0\right)\right\} \\ &= \nu_N^1(\pi_0^0, a) + \nu_N^2(\pi_0^0, a) \end{aligned}$$

with

$$\nu_N^1(\pi_0^0, a) = \sqrt{n}\begin{pmatrix} \frac{1}{n}\sum_{i=1}^{n}\left(Q^\tau_{\rho|(\pi^0,a_i)}(\pi_0^0, a_i) - \mathbb{E}Q^\tau_{\rho|(\pi^0,a)}(\pi_0^0, a)\right) \\ \frac{1}{n}\sum_{i=1}^{n} h_0(\dot{Q}^\tau_{\rho|(\pi^0,a_i)}(\pi_0^0, a_i) - \mathbb{E}\dot{Q}^\tau_{\rho|(\pi^0,a)}(\pi_0^0, a)) \end{pmatrix}$$

and

$$\nu_N^2(\pi_0^0, a) = \mathbb{E}\sqrt{H_{0,N}}\begin{pmatrix} \widehat{Q^\tau_{\rho|(\pi^0,a)}}(\pi_0^0, a) - Q^\tau_{\rho|(\pi^0,a)}(\pi_0^0, a) \\ h_0(\widehat{\dot{Q}^\tau_{\rho|(\pi^0,a)}}(\pi_0^0, a) - \dot{Q}^\tau_{\rho|(\pi^0,a)}(\pi_0^0, a)) \end{pmatrix}.$$

The latter does not depend on the first stage approximation due to the fact that $\sqrt{\frac{H_{0,N}}{H_N}} \to 0$ as $N \to \infty$ (dimensionality, see Section 1.4), such that the second term in Lemma 3 vanishes. Thus, for given a, the asymptotic distribution of $\widehat{\theta}_N$ can be derived analog to Lemma 5:

$$\sqrt{H_{0,N}}\begin{pmatrix} \widehat{Q^\tau_{\rho|(\pi^0,a)}}(\pi_0^0, a) - Q^\tau_{\rho|(\pi^0,a)}(\pi_0^0, a) \\ h_0(\widehat{\dot{Q}^\tau_{\rho|(\pi^0,a)}}(\pi_0^0, a) - \dot{Q}^\tau_{\rho|(\pi^0,a)}(\pi_0^0, a)) \end{pmatrix} \rightsquigarrow \mathcal{N}\left(\frac{h_0^2}{2}B_0^0(\pi_0, a), \frac{\tau(1-\tau)B_1^0}{f_{(\pi^0,a)}(\pi_0^0, a)f^2_{\rho|(\pi^0,a)}(Q^\tau_{\rho|(\pi^0,a)}(\pi_0^0, a))}\right)$$

with

$$B_{0j}^0(\pi_0^0, a) = \operatorname{tr}\left\{\ddot{Q}^\tau_{\rho|(\pi^0,a)}(\pi_0^0, a)\int uu^T\begin{pmatrix}1\\u\end{pmatrix}_j K(u)\,du\right\}$$

and B_1^0 as defined in Lemma 5, with dimensions K_0 instead of K. Then it follows immediately from consistency and the law of large numbers that the expectation with respect to the taste shocks $\mathbb{E}\nu_N^2 = \mathbb{E}[\nu_N^1 + \nu_N^2]$ is

$$\mathbb{B}(\pi_0^0) = \int \nu_N(\pi_0^0, a) f_a(a) da = \frac{h_0^2}{2} B_0^0 \int \ddot{Q}^{\tau}_{\rho|\pi^0}(\pi_0^0, a) f_a(a) da. \tag{1.24}$$

Using the law of total variance, the variance of ν_N^2 under $\mathbf{P}_0$ can be decomposed as

$$\begin{aligned}\mathbb{V}_{22}(\pi_0^0) &= \int \frac{\tau(1-\tau)B_1^0 f_a(a)da}{f_{(\pi^0,a)}(\pi_0^0,a) f^2_{\rho|(\pi^0,a)}(Q^{\tau}_{\rho|(\pi^0,a)}(\pi_0^0,a))} + \frac{h_0^4}{4}(B_0^0)^2 \int \left\{ \ddot{Q}^{\tau}_{\rho|\pi^0}(\pi_0^0,a) - \int \ddot{Q}^{\tau}_{\rho|\pi^0}(\pi_0^0,a) f_a(a) da \right\}^2 f_a(a) da \\ &= \int \frac{\tau(1-\tau)B_1^0 f_a(a)da}{f_{(\pi^0,a)}(\pi_0^0,a) f^2_{\rho|(\pi^0,a)}(Q^{\tau}_{\rho|(\pi^0,a)}(\pi_0^0,a))} + o_{P_0}(|h_0^4|)\end{aligned}$$

where the last term is of small order due to our finite second moments assumption on ε^s for all $s \in I_S$ and the regularity assumption on the individuals utilities (Assumption 1.3.(i)). The variance of ν_N^1 follows from the CLT and the delta rule

$$\mathbb{V}_{11}(\pi_0^0) = \mathbb{E}\begin{bmatrix} \dot{Q}^{\tau}_{\rho|(\pi^0,a)}(\pi_0^0,a)^T \Sigma_a \dot{Q}^{\tau}_{\rho|(\pi^0,a)}(\pi_0^0,a) & \dot{Q}^{\tau}_{\rho|(\pi^0,a)}(\pi_0^0,a)^T \Sigma_a \ddot{Q}^{\tau}_{\rho|(\pi^0,a)}(\pi_0^0,a) \\ \ddot{Q}^{\tau}_{\rho|(\pi^0,a)}(\pi_0^0,a)^T \Sigma_a \dot{Q}^{\tau}_{\rho|(\pi^0,a)}(\pi_0^0,a) & \ddot{Q}^{\tau}_{\rho|(\pi^0,a)}(\pi_0^0,a)^T \Sigma_a \ddot{Q}^{\tau}_{\rho|(\pi^0,a)}(\pi_0^0,a) \end{bmatrix}.$$

For the covariance we can proceed similarly

$$\begin{aligned}\mathbb{V}_{12}(\pi_0^0) =& \sqrt{\frac{\tau(1-\tau)}{h^K}} D_0(\pi_0^0)^{-1} \mathbb{E}\begin{pmatrix} 1 & h^{-1}(\pi^0-\pi_0^0)^T \\ h^{-1}(\pi^0-\pi_0^0) & h^{-2}(\pi^0-\pi_0^0)(\pi^0-\pi_0^0)^T \end{pmatrix}^{1/2} K\left(\frac{\pi^0-\pi_0^0}{h}\right) \\ &\times \left(\Sigma_a^{1/2}\left[\dot{Q}^{\tau}_{\rho|(\pi^0,a)}(\pi_0^0,a), \ddot{Q}^{\tau}_{\rho|(\pi^0,a)}(\pi_0^0,a)\right]\right)^T.\end{aligned}$$

Writing $\nu^1 + \nu^s = (\iota_2^T \otimes I_{K_0})\left[\nu^1, \nu^2\right]^T$ the asymptotic variance covariance matrix[13] of the local average structural derivative becomes

$$\mathbb{V}(\pi_0^0) = (\iota_2^T \otimes I_{K_0}) \begin{bmatrix} \mathbb{V}_{11}(\pi_0^0) & \mathbb{V}_{12}(\pi_0^0) \\ \mathbb{V}_{12}(\pi_0^0)^T & \mathbb{V}_{22}(\pi_0^0) \end{bmatrix} (I_{K_0} \otimes \iota_2). \tag{1.25}$$

□

Lemma 4 (Almost Ideal Demand System)**.** *Demands for public goods and the sharing rule are defined by*

$$\begin{aligned}\begin{pmatrix} x_0 \\ \rho \end{pmatrix} = \frac{w}{p^0} & \frac{b_2(p^2,\varepsilon^2)}{\overline{b}(p^1,p^2,w,z^{\mu},\varepsilon^1,\varepsilon^2,\varepsilon^{\mu}) + b_1(p^1,\varepsilon^1)b_2(p^2,\varepsilon^2)\overline{\eta}(p^1,p^2,w,z^{\mu},\varepsilon^0,\varepsilon^{\mu})} \\ & \times \begin{pmatrix} b_1(p^1,\varepsilon^1)\overline{\eta}(p^1,p^2,w,z^{\mu},\varepsilon^0,\varepsilon^{\mu}) \\ p^0\mu(p^1,p^2,w,z^{\mu},\varepsilon^{\mu}) \end{pmatrix}\end{aligned} \tag{1.10}$$

which are functions of both private and public taste shocks where $\overline{\eta} = \mu\eta_1 + (1-\mu)\eta_2 + \varepsilon^0$ *and* $\overline{b} = \mu b_1(\varepsilon^1) + (1-\mu)b_2(\varepsilon^2)$.

[13] Note that while the asymptotic variance covariance matrix depends on the second derivative of the conditional quantile of ρ, we can estimate this term by considering the asymptotic variance of the sample mean $n^{-1}\sum \widehat{Q}(\pi_0^0, a_i^*)$ for given π_0^0 with respect to $a_i^* \sim \widehat{F}_{\hat{a}}$

Proof. In the first stage, the two member household we specified optimizes

$$\max_{x^0,\rho} \mu(w,p^1,p^2,z^\mu,\varepsilon^\mu)\left[\frac{\log\rho^1-\log a_1(p^1)}{b_1(p^1,\varepsilon^1)}+(\eta_1+\varepsilon^0)\log x^0\right]$$
$$+\left(1-\mu(w,p^1,p^2,z^\mu,\varepsilon^\mu)\right)\left[\frac{\log(\rho^2,-\log a_2(p^2)}{b_2(p^2,\varepsilon^2)}+(\eta_2+\varepsilon^0)\log x^0\right].$$

subject to $w=\rho^1-p^0x^0$

First order conditions with respect to x^0 are ρ^1 are, respectively,

$$\mu(w,p^1,p^2,z^\mu,\varepsilon^\mu)\frac{(\eta_1+\varepsilon^0)}{x^0}+\left(1-\mu(w,p^1,p^2,z^\mu,\varepsilon^\mu)\right)\frac{(\eta_2+\varepsilon^0)}{x^0}=m^0p^0 \tag{1.26}$$

and

$$\mu(w,p^1,p^2,z^\mu,\varepsilon^\mu)\frac{1}{\rho^1b_1(p^1,\varepsilon^1)}=\left(1-\mu(w,p^1,p^2,z^\mu,\varepsilon^\mu)\right)\frac{1}{\rho^2b_2(p^2,\varepsilon^2)}=m^0 \tag{1.27}$$

where m^0 represents the Lagrange multiplier arising from the budget constraint. Using the the two equations in (1.27) together with the budget constraint we get

$$\rho^1=(w-\rho^1-p^0x^0)\frac{\mu(w,p^1,p^2,z^\mu,\varepsilon^\mu)}{1-\mu(w,p^1,p^2,z^\mu,\varepsilon^\mu)}\frac{b_2(p^2,\varepsilon^2)}{b_1(p^1,\varepsilon^1)}$$

which we can simplify to

$$\rho^1=\frac{\mu(w,p^1,p^2,z^\mu,\varepsilon^\mu)b_2(p^2,\varepsilon^2)}{(1-\mu(w,p^1,p^2,z^\mu,\varepsilon^\mu))\,b_1(p^1,\varepsilon^1)+\mu(w,p^1,p^2,z^\mu,\varepsilon^\mu)b_2(p^2,\varepsilon^2)}\left(w-p^0x^0\right). \tag{1.28}$$

Substituting the Lagrange multiplier m^0 in equation (1.26) using the first equation in (1.27) we get

$$\mu(w,p^1,p^2,z^\mu,\varepsilon^\mu)\frac{(\eta_1+\varepsilon^0)}{x^0}+\left(1-\mu(w,p^1,p^2,z^\mu,\varepsilon^\mu)\right)\frac{(\eta_2+\varepsilon^0)}{x^0}=p^0\frac{\mu(w,p^1,p^2,z^\mu,\varepsilon^\mu)}{\rho^1b_1(p^1,\varepsilon^1)}.$$

By substituting ρ^1 using equation (1.28) an defining for notational simplicity:

$$\overline{\eta}(w,p^1,p^2,z^\mu,\varepsilon^\mu,\varepsilon^0)=\mu(w,p^1,p^2,z^\mu,\varepsilon^\mu)\eta_1+(1-\mu(w,p^1,p^2,z^\mu,\varepsilon^\mu))\eta_2+\varepsilon^0$$
$$\overline{b}(w,p^1,p^2,z^\mu,\varepsilon^\mu)=\mu(w,p^1,p^2,z^\mu,\varepsilon^\mu)b_1(\varepsilon^1)+(1-\mu(w,p^1,p^2,z^\mu,\varepsilon^\mu))b_2(\varepsilon^2)$$

we get after simplifying

$$\left(w-p^0x^0\right)\overline{\eta}(w,p^1,p^2,z^\mu,\varepsilon^\mu,\varepsilon^0)=p^0x^0\frac{\overline{b}(w,p^1,p^2,z^\mu,\varepsilon^\mu)}{b_1(p^1,\varepsilon^1)b_2(p^2,\varepsilon^2)}$$

and hence public consumption in quasi Cobb-Douglas form

$$x^0=\frac{b_1(p^1,\varepsilon^1)b_2(p^2,\varepsilon^2)\overline{\eta}(w,p^1,p^2,z^\mu,\varepsilon^\mu,\varepsilon^0)}{\overline{b}(w,p^1,p^2,z^\mu,\varepsilon^\mu)+b_1(p^1,\varepsilon^1)b_2(p^2,\varepsilon^2)\overline{\eta}(w,p^1,p^2,z^\mu,\varepsilon^\mu,\varepsilon^0)}\frac{w}{p^0}$$

with parameter that is a combination of the members' tastes for public goods and private goods, where the latter enters through the trade-off between public and private goods. Similarly, by using equation (1.28) again, we get the sharing rule

$$\rho^1 = \frac{w\mu(w, p^1, p^2, z^\mu, \varepsilon^\mu) b_2(p^2, \varepsilon^2)}{\overline{b}(w, p^1, p^2, z^\mu, \varepsilon^\mu) + b_1(p^1, \varepsilon^1) b_2(p^2, \varepsilon^2) \overline{\eta}(w, p^1, p^2, z^\mu, \varepsilon^\mu, \varepsilon^0)}$$

which is a (linear) function of the Pareto weight for the respective individual. □

Further Application Results

	$\tau = 0.1$			$\tau = 0.25$			$\tau = 0.5$			$\tau = 0.75$			$\tau = 0.9$		
(p^h, ρ^h)	$\widehat{\epsilon}$	$\widehat{\sigma_\epsilon}$		$\widehat{\epsilon}$	$\widehat{\sigma_\epsilon}$		$\widehat{\epsilon}$	$\widehat{\sigma_\epsilon}$		$\widehat{\epsilon}$	$\widehat{\sigma_\epsilon}$		$\widehat{\epsilon}$	$\widehat{\sigma_\epsilon}$	
(7.7 , 853.3)	-0.263	0.103	**	-0.344	0.078	***	-0.379	0.046	***	-0.409	0.052	***	-0.375	0.061	***
(7.7 , 1132.1)	-0.212	0.049	***	-0.270	0.033	***	-0.309	0.113	***	-0.346	0.198	*	-0.327	0.461	
(7.7 , 1455.6)	-0.154	0.048	***	-0.196	0.058	***	-0.246	0.316		-0.290	444.389		-0.278	-	
(10.2 , 853.3)	-0.421	0.296		-0.533	0.381		-0.599	0.136	***	-0.634	0.153	***	-0.573	0.048	***
(10.2 , 1132.1)	-0.332	0.112	***	-0.427	0.062	***	-0.480	0.057	***	-0.525	0.094	***	-0.486	0.110	***
(10.2 , 1455.6)	-0.256	0.054	***	-0.315	0.056	***	-0.377	0.168	**	-0.431	0.257	*	-0.407	0.814	
(12.9 , 853.3)	-0.655	0.969		-0.836	3.904		-0.943	1.630		-0.986	0.567	*	-0.841	0.146	***
(12.9 , 1132.1)	-0.517	0.469		-0.641	0.251	**	-0.733	0.209	***	-0.791	0.137	***	-0.707	0.064	***
(12.9 , 1455.6)	-0.375	0.104	***	-0.501	0.058	***	-0.569	0.067	***	-0.637	0.132	***	-0.597	0.197	***

Table 1.7: Elasticity of x_1^h with respect to p^h

	$\tau = 0.1$			$\tau = 0.25$			$\tau = 0.5$			$\tau = 0.75$			$\tau = 0.9$		
(p^h, ρ^h)	$\widehat{\epsilon}$	$\widehat{\sigma_\epsilon}$		$\widehat{\epsilon}$	$\widehat{\sigma_\epsilon}$		$\widehat{\epsilon}$	$\widehat{\sigma_\epsilon}$		$\widehat{\epsilon}$	$\widehat{\sigma_\epsilon}$		$\widehat{\epsilon}$	$\widehat{\sigma_\epsilon}$	
(7.7 , 853.3)	0.434	0.231	*	0.464	0.125	***	0.469	0.059	***	0.468	0.057	***	0.410	0.063	***
(7.7 , 1132.1)	0.470	0.098	***	0.486	0.052	***	0.509	0.156	***	0.525	0.271	*	0.474	0.630	
(7.7 , 1455.6)	0.458	0.099	***	0.465	0.105	***	0.524	0.557		0.567	781.123		0.518	-	
(10.2 , 853.3)	0.504	0.499		0.547	0.403		0.563	0.162	***	0.553	0.143	***	0.476	0.040	***
(10.2 , 1132.1)	0.537	0.198	***	0.578	0.080	***	0.597	0.065	***	0.604	0.100	***	0.534	0.115	***
(10.2 , 1455.6)	0.549	0.093	***	0.559	0.080	***	0.606	0.228	***	0.639	0.344	*	0.579	1.089	
(12.9 , 853.3)	0.597	1.184		0.676	2.483		0.698	1.373		0.680	0.418		0.556	0.123	***
(12.9 , 1132.1)	0.630	0.483		0.687	0.305	**	0.720	0.234	***	0.722	0.122	***	0.616	0.056	***
(12.9 , 1455.6)	0.621	0.153	***	0.683	0.074	***	0.716	0.077	***	0.743	0.142	***	0.666	0.208	***

Table 1.8: Elasticity of x_1^h with respect to ρ^h

	$\tau = 0.1$			$\tau = 0.25$			$\tau = 0.5$			$\tau = 0.75$			$\tau = 0.9$		
(p^w, ρ^w)	$\widehat{\epsilon}$	$\widehat{\sigma_\epsilon}$		$\widehat{\epsilon}$	$\widehat{\sigma_\epsilon}$		$\widehat{\epsilon}$	$\widehat{\sigma_\epsilon}$		$\widehat{\epsilon}$	$\widehat{\sigma_\epsilon}$		$\widehat{\epsilon}$	$\widehat{\sigma_\epsilon}$	
(9.4 , 843.8)	-0.127	0.084		-0.197	0.058	***	-0.290	0.047	***	-0.356	0.060	***	-0.373	0.129	***
(9.4 , 1076.2)	-0.096	0.096		-0.140	0.051	***	-0.225	0.076	***	-0.299	0.489		-0.329	0.571	
(9.4 , 1378.2)	-0.054	0.112		-0.070	0.067		-0.144	0.169		-0.233	2.576		-0.284	2.950	
(11.3 , 843.8)	-0.188	0.204		-0.299	0.228		-0.401	0.101	***	-0.479	0.084	***	-0.477	0.052	***
(11.3 , 1076.2)	-0.143	0.106		-0.227	0.078	***	-0.327	0.069	***	-0.405	0.287		-0.423	0.440	
(11.3 , 1378.2)	-0.095	0.094		-0.138	0.067	**	-0.234	0.152		-0.322	1.389		-0.370	3.041	
(14.3 , 843.8)	-0.332	3.644		-0.493	0.510		-0.630	1.759		-0.717	0.537		-0.673	0.125	***
(14.3 , 1076.2)	-0.251	0.241		-0.389	0.193	**	-0.519	0.146	***	-0.608	0.105	***	-0.590	0.082	***
(14.3 , 1378.2)	-0.175	0.092	*	-0.277	0.070	***	-0.403	0.104	***	-0.494	0.368		-0.512	1.617	

Table 1.9: Elasticity of x_1^w with respect to p^w

	$\tau = 0.1$			$\tau = 0.25$			$\tau = 0.5$			$\tau = 0.75$			$\tau = 0.9$		
(p^w, ρ^w)	$\widehat{\epsilon}$	$\widehat{\sigma_\epsilon}$		$\widehat{\epsilon}$	$\widehat{\sigma_\epsilon}$		$\widehat{\epsilon}$	$\widehat{\sigma_\epsilon}$		$\widehat{\epsilon}$	$\widehat{\sigma_\epsilon}$		$\widehat{\epsilon}$	$\widehat{\sigma_\epsilon}$	
(9.4 , 843.8)	0.301	0.198		0.324	0.145	**	0.370	0.067	***	0.414	0.054	***	0.412	0.099	***
(9.4 , 1076.2)	0.327	0.171	*	0.336	0.070	***	0.384	0.080	***	0.448	0.473		0.459	0.545	
(9.4 , 1378.2)	0.339	0.193	*	0.321	0.096	***	0.364	0.216	*	0.459	3.225		0.504	3.626	
(11.3 , 843.8)	0.330	0.701		0.375	0.404		0.416	0.140	***	0.460	0.091	***	0.445	0.049	***
(11.3 , 1076.2)	0.354	0.191	*	0.391	0.093	***	0.442	0.067	***	0.498	0.233	**	0.496	0.350	
(11.3 , 1378.2)	0.369	0.154	**	0.376	0.084	***	0.435	0.166	***	0.514	1.439		0.548	3.115	
(14.3 , 843.8)	0.392	6.177		0.455	0.818		0.507	0.983		0.540	0.413		0.509	0.115	***
(14.3 , 1076.2)	0.414	0.304		0.473	0.218	**	0.536	0.110	***	0.584	0.091	***	0.558	0.057	***
(14.3 , 1378.2)	0.426	0.124	***	0.469	0.080	***	0.544	0.099	***	0.610	0.304	**	0.607	1.338	

Table 1.10: Elasticity of x_1^w with respect to ρ^w

	$\tau = 0.1$		$\tau = 0.25$		$\tau = 0.5$		$\tau = 0.75$		$\tau = 0.9$	
$(p^w/p^h, w)$	$\widehat{\epsilon}$	$\widehat{\sigma_\epsilon}$	$\widehat{\epsilon}$	$\widehat{\sigma_\epsilon}$	$\widehat{\epsilon}$	$\widehat{\sigma_\epsilon}$	$\widehat{\epsilon}$	$\widehat{\sigma_\epsilon}$	$\widehat{\epsilon}$	$\widehat{\sigma_\epsilon}$
(0.6 , 3564.1)	0.035	0.294	-0.026	0.070	-0.078	0.061	-0.044	0.043	-0.074	0.117
(0.6 , 4337.1)	0.030	0.150	-0.021	0.059	-0.064	0.056	-0.036	0.053	-0.061	0.193
(0.6 , 5225.5)	0.024	0.072	-0.017	0.056	-0.051	0.048	-0.030	0.101	-0.051	0.345
(0.9 , 3564.1)	0.039	0.391	-0.038	0.093	-0.119	0.079	-0.064	0.056	-0.111	0.142
(0.9 , 4337.1)	0.032	0.178	-0.030	0.073	-0.093	0.072	-0.053	0.063	-0.091	0.199
(0.9 , 5225.5)	0.029	0.087	-0.023	0.071	-0.076	0.061	-0.043	0.130	-0.075	0.398
(1.2 , 3564.1)	0.042	0.587	-0.054	0.140	-0.162	0.118	-0.085	0.080	-0.149	0.190
(1.2 , 4337.1)	0.034	0.283	-0.042	0.107	-0.128	0.109	-0.068	0.101	-0.120	0.267
(1.2 , 5225.5)	0.029	0.125	-0.033	0.103	-0.102	0.090	-0.056	0.217	-0.099	0.663

Table 1.11: Elasticity of c_c^0 with respect to p^w/p^h

	$\tau = 0.1$			$\tau = 0.25$			$\tau = 0.5$			$\tau = 0.75$			$\tau = 0.9$		
$(p^w/p^h, w)$	$\widehat{\epsilon}$	$\widehat{\sigma_\epsilon}$		$\widehat{\epsilon}$	$\widehat{\sigma_\epsilon}$		$\widehat{\epsilon}$	$\widehat{\sigma_\epsilon}$		$\widehat{\epsilon}$	$\widehat{\sigma_\epsilon}$		$\widehat{\epsilon}$	$\widehat{\sigma_\epsilon}$	
(0.6 , 3564.1)	1.499	1.324		1.031	0.348	***	0.860	0.236	***	0.910	0.217	***	0.923	0.567	
(0.6 , 4337.1)	1.356	0.757	*	1.013	0.277	***	0.895	0.211	***	0.931	0.253	***	0.939	0.531	*
(0.6 , 5225.5)	1.272	0.511	**	1.007	0.315	***	0.905	0.251	***	0.945	0.566	*	0.946	1.418	
(0.9 , 3564.1)	1.525	1.244		1.106	0.350	***	0.916	0.251	***	0.934	0.222	***	0.963	0.600	
(0.9 , 4337.1)	1.355	0.638	**	1.080	0.269	***	0.926	0.224	***	0.963	0.237	***	0.980	0.448	**
(0.9 , 5225.5)	1.281	0.448	***	1.057	0.312	***	0.946	0.246	***	0.963	0.532	*	0.971	1.047	
(1.2 , 3564.1)	1.493	1.246		1.130	0.358	***	0.967	0.266	***	0.968	0.219	***	1.001	0.749	
(1.2 , 4337.1)	1.360	0.662	**	1.097	0.283	***	0.974	0.238	***	0.964	0.252	***	1.001	0.453	**
(1.2 , 5225.5)	1.294	0.437	***	1.080	0.321	***	0.966	0.256	***	0.976	0.554	*	0.993	1.100	

Table 1.12: Elasticity of c_c^0 with respect to w

$(p^w/p^h, w)$	$\tau = 0.1$			$\tau = 0.25$			$\tau = 0.5$			$\tau = 0.75$			$\tau = 0.9$		
	$\hat{\epsilon}$	$\widehat{\sigma_\epsilon}$		$\hat{\epsilon}$	$\widehat{\sigma_\epsilon}$		$\hat{\epsilon}$	$\widehat{\sigma_\epsilon}$		$\hat{\epsilon}$	$\widehat{\sigma_\epsilon}$		$\hat{\epsilon}$	$\widehat{\sigma_\epsilon}$	
(0.6 , 3564.1)	0.521	0.286	*	0.477	0.158	***	0.462	0.095	***	0.419	0.059	***	0.374	0.043	***
(0.6 , 4337.1)	0.436	0.141	***	0.404	0.083	***	0.396	0.052	***	0.363	0.059	***	0.322	0.045	***
(0.6 , 5225.5)	0.369	0.077	***	0.343	0.053	***	0.343	0.060	***	0.316	0.059	***	0.277	0.081	***
(0.9 , 3564.1)	0.614	0.173	***	0.564	0.103	***	0.556	0.075	***	0.512	0.076	***	0.466	0.057	***
(0.9 , 4337.1)	0.534	0.102	***	0.491	0.066	***	0.491	0.064	***	0.456	0.068	***	0.412	0.076	***
(0.9 , 5225.5)	0.458	0.058	***	0.428	0.048	***	0.432	0.073	***	0.403	0.094	***	0.363	0.072	***
(1.2 , 3564.1)	0.679	0.157	***	0.627	0.101	***	0.618	0.086	***	0.582	0.094	***	0.531	0.089	***
(1.2 , 4337.1)	0.594	0.091	***	0.556	0.068	***	0.555	0.103	***	0.520	0.119	***	0.478	0.107	***
(1.2 , 5225.5)	0.521	0.060	***	0.491	0.074	***	0.500	0.111	***	0.467	0.128	***	0.426	0.160	***

Table 1.13: Elasticity of ρ^w with respect to p^w/p^h

$(p^w/p^h, w)$	$\tau = 0.1$			$\tau = 0.25$			$\tau = 0.5$			$\tau = 0.75$			$\tau = 0.9$		
	$\hat{\epsilon}$	$\widehat{\sigma_\epsilon}$		$\hat{\epsilon}$	$\widehat{\sigma_\epsilon}$		$\hat{\epsilon}$	$\widehat{\sigma_\epsilon}$		$\hat{\epsilon}$	$\widehat{\sigma_\epsilon}$		$\hat{\epsilon}$	$\widehat{\sigma_\epsilon}$	
(0.6 , 3564.1)	0.899	0.665		0.848	0.391	**	0.780	0.273	***	0.721	0.150	***	0.730	0.082	***
(0.6 , 4337.1)	0.907	0.344	***	0.869	0.257	***	0.811	0.168	***	0.757	0.121	***	0.766	0.092	***
(0.6 , 5225.5)	0.918	0.267	***	0.882	0.202	***	0.843	0.175	***	0.789	0.153	***	0.796	0.242	***
(0.9 , 3564.1)	0.734	0.311	**	0.693	0.236	***	0.635	0.161	***	0.614	0.109	***	0.620	0.074	***
(0.9 , 4337.1)	0.774	0.208	***	0.730	0.156	***	0.681	0.123	***	0.662	0.101	***	0.665	0.114	***
(0.9 , 5225.5)	0.796	0.157	***	0.762	0.146	***	0.720	0.143	***	0.701	0.182	***	0.707	0.260	***
(1.2 , 3564.1)	0.638	0.212	***	0.593	0.155	***	0.537	0.114	***	0.539	0.085	***	0.541	0.079	***
(1.2 , 4337.1)	0.675	0.135	***	0.635	0.111	***	0.585	0.103	***	0.586	0.110	***	0.591	0.154	***
(1.2 , 5225.5)	0.703	0.113	***	0.672	0.124	***	0.633	0.145	***	0.630	0.222	***	0.636	0.412	

Table 1.14: Elasticity of ρ^w with respect to w

Time-use and Consumption Categories

Public consumption

(i) mortgage: interest plus amortization (gross)
(ii) rent (NOT including costs of gas and electricity)
(iii) general utilities (heating, electricity, water, telephone, Internet, etc)
(iv) transport and means of transport (public transport; own car: gasoline/diesel and maintenance)
(v) insurances (home insurance, car insurance, health insurance, etc.)
(vi) children's daycare (day care center, out-of-school supervision, guest parents, homework guidance)
(vii) alimony and financial support for children not (or no longer) living at home
(viii) debts and loans (excluding mortgage)
(ix) daytrips and holidays with the whole family or part of the family (flight tickets, hotel, restaurant bills for the family, etc.)
(x) expenditure on cleaning the house or maintaining the garden
(xi) eating at home (food, drinks, candy, etc.)
(xii) other

Private consumption

(i) food and drinks outside the house (restaurant, cafe, company canteen, etc., but NO restaurant bills for the family)

(ii) cigarettes and other tobacco products

(iii) clothing (clothes, shoes, jewelry, etc.)

(iv) personal care products and services (hair care, body care, hairdresser, manicure, etc.)

(v) medical care and health costs NOT covered by insurance (medicines, doctor, dentist, hospital bills, maternity care, spectacles, hearing aids, etc.)

(vi) leisure time expenditure (film, theater, hobbies, sports activities, photography, books, CDs/DVDs, expenditure during daytrips or travel without family, etc.)

(vii) (further) schooling (expenditure on courses, enrolment fees, etc.)

(viii) gifts and presents (for family, friends, charity, etc.)

(ix) other

CHAPTER 2

The Collective Axiom in a Heterogeneous Population

Introduction

It is often argued that the traditional unitary household consumption model, which assumes that a household consists of only one (aggregate) decision maker, is insufficient to describe household behaviour. In addition to this, many policy relevant questions concerning family economics, e.g. pooling of taxable income or to whom to provide childcare benefits, cannot be answered within a unitary setting. From an empirical point of view some authors find evidence that the symmetry property of the Slutsky substitution matrix implied by the unitary model should be rejected [Browning & Chiappori, 1998; Cherchye et al., 2009]. A widely accepted alternative to the unitary model is known as the collective household consumption model [Chiappori, 1988, 1992], in which household members bargain over their consumption choices and are assumed to reach a Pareto efficient outcome. This model implies a Slutsky matrix that can be decomposed into a regular negative semi-definite, symmetric term and a rank deficient outer product which is commonly referred to as the SNR$(S-1)$ condition and thus provides a theoretical foundation for the lacking symmetry property that is often found in the data [Chiappori & Ekeland, 2006, 2009].

There are two strands in the literature of testing the collective household model versus an alternative one. The first type considers a continuous demand system, based on which the Slutsky matrix is constructed and its rank is tested [Browning & Chiappori, 1998]. The second approach is fully non-parametric and based on revealed preference restrictions [Cherchye et al., 2007, 2009]. While both these strands have their merits, the first one has the drawback that it is very difficult to allow for unobserved heterogeneity either with respect to preferences or with respect to household bargaining, even in a fully parametric setting. The revealed preference approach on the other hand allows for unobserved heterogeneity as long as panel data is available. Both these tests are based on aggregate household consumption data which is feasible but, in the latter case, comes with the drawback that it may not have enough power and will thus often fail to reject the collective model since households are considered to be consistent with the revealed preference axiom if there exists a hypothetical within-household demand allocation satisfying the axiom. In order to test whether or not a household is *actually* rational in a collective sense, researchers often use additional consumption data from single households to [Browning et al., 2013; Dunbar et al., 2013b]. This however requires the assumption of stable preferences, where stability is to be interpreted over different household compositions. In other words, the transition from being single to being in a couple does not affect individual consumption preferences.

In this paper we construct a test of the validity of this stable preference assumption which is fully nonparametric and allows for a heterogeneous population, i.e. unobserved heterogeneity both with respect to preferences and bargaining power. In order to achieve this, we will construct discrete household types for both couples and singles. We will assume that we fully observe the marginal distribution of their respective continuous consumption choices in at least three different price regimes, which we will then map into the discrete choice space. For this we require panel data. The choice space will be constructed in a way that ensures that two households for which the same revealed preference inequalities hold form an equivalence class. In order to check rationality of a household in a first step we define hypothetical household types, which are characterized by the discrete choice of a couple's household and the choices of both husband and wife within that household. Since the latter two are not observed we will make use of the stable preference assumption and a separability restriction that allow us to treat single individuals as if they were in a couple and obtain their preference relations from single individuals. A household type is then characterized as the three-tuple of a couple's aggregate household choice, husband's choice and wife's choice. The fully characterized type space is defined as all possible combinations thereof. In a second step we will then link observed choices of both singles and couples to the space of fully characterized household types. Since we do not observe the joint distribution supported on the constructed hypothetical type space, but only marginal distributions of consumption choices for single females, single males and couples, respectively, we will make use of a stochastic revealed preference setting [McFadden & Richter, 1991; McFadden, 2005], which leads to a partial identification approach of the joint distribution or copula. The inference of this test is then based on the ideas of Hoderlein & Stoye [2014]; Kitamura & Stoye [2013]. Under the hypothesis of stable preferences, we can then for each of these hypothetical types, not only determine whether or not for a given type of household there exists a feasible consumption allocation which makes this household rational, using the necessary conditions in Cherchye et al. [2007], but also whether this is actually the case for this type using the preference relations obtained from singles. The test is then constructed in a way that we consider households that satisfy the necessary conditions as a baseline case, and then compare it with the case in which we have added the information of singles. The difference between these two sets are the households for which the stable preference assumption does not hold.

We will now briefly discuss the existing approaches known in the literature how to test the validity of the collective model and how the contribution of this paper relates to them. Testing the collective model using the Slutsky matrix requires estimating household demands and derivatives thereof. Browning & Chiappori [1998] construct a test of collective rationality based on a parametric almost ideal demand system with additive (measurement) errors. While an almost ideal demand system provides a locally flexible functional parametric form of demands which can be readily tested for the SNR$(S-1)$ condition, it is prone to potential misspecification error. In addition to this, identification of continuous demand systems in the presence of general unobserved heterogeneity is difficult due to the rather complex structure of the collective model, which leads to demands that are non-separable with respect to the random preference and bargaining parameters representing heterogeneous consumers. Using observable intra-household allocation of consumption Hubner [2015] derives restrictions on utility functions and Pareto weights that ensure nonparametric identification of vector-valued and continuous household demand functions using a global invertibility argument, which could be used to non-parametrically test the SNR$(S-1)$ condition in the presence of unobserved heterogeneity. In the context of a continuous demand setting, the use of single household data is not new, although it is usually not used for testing but rather for the recovery of the so-called (conditional) sharing rule, which determines the division of the endowment among the members within the household. Such an approach was used for example in a collective labour supply setting by Barmby & Smith [2001] and Vermeulen et al. [2006], and introduced to a collective consumption setting by Browning et al. [2013] who show how single data can be used to identify equivalence scales in a homogeneous household setting. Dunbar et al. [2013b] extend this approach to a setting with unobserved heterogeneity by considering completely random sharing rules and show that under certain preference restrictions and the existence of a number of so-called distribution factors, the joint distribution of sharing rule levels is identified.

To the knowledge of the author, the use of singles data in the context of the fully non-parametric global way to model collective households using revealed preference restrictions [Cherchye et al., 2007, 2009] is novel. The extensive use of the stable preference assumption in the literature motivates a test of this hypothesis. A revealed preference based approach is desirable since it allows us to use a fully stochastic random utility and random bargaining power version of the collective model, without requiring assumptions on how unobserved heterogeneity enters the models' primitives, and thus circumvents the identification problems of the local, continuous approach. On the other hand, the discrete nature of this approach makes it difficult to work with single data, since the model in its most general form involves externalities modeled by Lindahl prices which implies that budget planes are individual-specific in a heterogeneous population and thus a finite-dimensional classification of types would not be feasible. Thus we consider a generalization of a Beckerian caring model [Becker, 1981]. This allows us to use the same choice space for both singles and couples. In order to discretize choices for singles and couples, we make use of panel data and a time-stability of preferences assumption. There exists a range of consumption datasets that would be feasible for our approach including the *Russian Longitudinal Monitoring Survey (RLMS)* and the *Spanish Continuous Family Expenditure Survey (Encuesta Continua de Presupuestos Familiares, ECPF)* which were used by Cherchye et al. [2011] and Adams et al. [2014] for similar purposes. This allows us to characterize the marginal distributions of consumption choices of both singles and couples.

In order to link the observed marginal distributions defined over choice types of singles and couples, to the joint distribution fully characterizing household behaviour we make use of the approaches by Hoderlein & Stoye [2014] and Kitamura & Stoye [2013]. Hoderlein & Stoye [2014] consider the weak axiom of revealed preference in the unitary model. In particular they use the fact that demands of a heterogeneous population observed in a given price regime can be characterized as random variables supported on a normalized budget set. Observing the same population in different price regimes (repeated cross-sections), one can then use copula techniques to derive (Frechet–Hoeffding) bounds on the probability that the population behaves irrationally, i.e. is not in line with the weak axiom. Kitamura & Stoye [2013] go one step further and fully discretize budget sets using a partition containing all the information that is relevant to test the strong axiom of revealed preference.

This paper is structured as follows. Section 2.2 specifies the collective model and discusses stable preferences and separability assumptions that will be necessary for our approach. Section 2.3 shows how the continuous consumption decision can be discretized for a household and each of its individuals without losing relevant information to test the collective axiom. Section 2.4 will discuss how we can construct a test statistic using the constructed type space and deals with the question of computational complexity. Section 2.5 provides simulation results using a three-good economy with three price regimes.

Specification

In this section we specify the standard collective model with two-person households and consumption externalities and show how we can make use of single households under the stable preference assumption. Let households, price regimes and goods be indexed by $i \in I_N$, $t \in I_T$, and $l \in I_L$, respectively[1]. Further let $\tilde{x}^c_{i,t} \in R^{LT}_+$ be continuous household consumption, which household $i \in I_N$ chooses optimizing its collective utility $u^m(\tilde{x}^m, \tilde{x}^f) + \mu(p)u^f(\tilde{x}^f, \tilde{x}^m)$ subject to budget $B_t = \{\tilde{x} \mid p_t\tilde{x} \leqslant 1\}$ and $\tilde{x}^c = \tilde{x}^f + \tilde{x}^m$, where $\tilde{x}^r$ is continuous individual private consumption of the respective spouses $r \in \{m, f\}$. Total household endowment is normalized to one. As such the expenditure on a given good category $l \in I_L$ denoted by $p_{t,l}\tilde{x}^c_{i,t,l}$ (e.g. food, transportation or electronics) represents its budget share, i.e. the fraction of the budget of a given period $t \in I_T$ that is allocated to purchase this good by household $i \in I_N$.

The collective model as defined above imposes restrictions on how both household and individual demands must change with respect to relative price changes. For self-sufficiency of this paper we state the axiom consisting of a set of necessary conditions below. A more detailed explanation and proof can be found in Cherchye et al. [2007].

[1]The set $I_J = \{1, \ldots, J\}$ refers to the index set running from 1 to J.

Definition 2.1 (Collective Axiom of Revealed Preference, Cherchye et al. [2007]). Suppose that there exists a pair of utility functions u^f and u^m that provide a collective rationalization of the set of observations $\left\{(p_t; \tilde{x}_t^c, \tilde{x}_t^f, \tilde{x}_t^m) : \tilde{x}_t^c = \tilde{x}_t^f + \tilde{x}_t^m, t \in I_T\right\}$. Then there exist preference relations[2] R_0^r and R^r for each $r \in \{c, m, f\}$ such that:

(i) if $\tilde{x}_s R_0^c \tilde{x}_t$, then $\tilde{x}_s R_0^f \tilde{x}_t$ or $\tilde{x}_s R_0^m \tilde{x}_t$

(ii) if $\tilde{x}_s R_0^r \tilde{x}_{s_1}, \tilde{x}_{s_1} R_0^r \tilde{x}_{s_2}, \ldots, \tilde{x}_{s_S} R_0^r \tilde{x}_t$ then $\tilde{x}_s R^r \tilde{x}_t$ for $r \in \{m, f\}$

(iii) if $\tilde{x}_s R_0^c \tilde{x}_t$ and $\tilde{x}_t R^r \tilde{x}_s$, then $\tilde{x}_s R_0^{r'} \tilde{x}_t$ for $r \neq r'$ where $r, r' \in \{m, f\}$

(iv) if $\tilde{x}_s R_0^c (\tilde{x}_{t_1} + \tilde{x}_{t_2})$ and $\tilde{x}_{t_1} R^r \tilde{x}_s$ then $\tilde{x}_s R_0^{r'} \tilde{x}_{t_2}$ for $r \neq r'$ where $r, r' \in \{m, f\}$.

(v) if $\tilde{x}_{s_1} R^f \tilde{x}_t$ and $\tilde{x}_{s_2} R^m \tilde{x}_t$ then $\neg (\tilde{x}_t R_0^c (\tilde{x}_{s_1} + \tilde{x}_{s_2}))$

(vi) if $\tilde{x}_s R^f \tilde{x}_t$ and $\tilde{x}_s R^m \tilde{x}_t$, then $\neg (\tilde{x}_t R_0^c \tilde{x}_s)$

where R^r is defined as $\tilde{x}_s R_0^r \tilde{x}_t$ whenever $p_s \tilde{x}_s^r \geqslant p_s \tilde{x}_t^r$ and R^r is the transitive closure of R_0^r [Afriat, 1967; Varian, 1982].

As the weak, strong and generalized axiom in the unitary model, the collective axiom considers demands in different price regimes. In addition to individual rationality of each spouse the collective model also requires Pareto efficiency, which together with monotonicity of preferences implies that every solution $\tilde{x}_{i,t}^c$ must lie on the budget set B_t for each $t \in I_T$. In a heterogeneous population, demands in each price regime are thus scattered on the respective budget planes. Since empirically not every household will have the same endowment, we will project heterogeneous demands onto normalized budget planes instead. This will be discussed in more detail in the empirical section below.

Demands of each spouse are in general not observable and thus the preference relations R^f and R^m are not identified directly from couples data. In the original approach [Cherchye et al., 2007, 2011] a household is considered rational if there exists a feasible consumption allocation between the spouses, or in other words a pair of hypothetical preference relations R^f and R^m, that is consistent with the collection axiom. Only households for which there does not exist such a pair of individual preference relations are considered irrational, which reduces the power of this test and may thus often accept the hypothesis that households behave according to the collective

[2]Note that R_0^c is just notation and not actually a preference relation, since household consumption is only a result of individual preferences.

model when it is not true[3]. Thus, we will take a different route here by exploiting information of single households in order to identify the preference relations R^f and R^m. For this let us consider a single household who chooses an optimal consumption bundle for a given period $t \in I_T$ by maximizing $u^r(\tilde{x}^r, \tilde{x}^{r'})$ subject to the constraint $\tilde{x}^r \in B_t$. Obviously, for single individuals the spouse's consumption $\tilde{x}^{r'}$ will be zero. Thus, in order to be able to model singles in a way that makes them informative for a spouses consumption behaviour we have to make a separability assumption. Let the $(L-1)$-dimensional vector of marginal rates of substitutions for $r \in \{m, f\}$ be denoted as $\mathrm{MRS}^r(x^r, x^{r'})$ with components $\mathrm{MRS}^r_l(x^r, x^{r'}) = \frac{\partial u^r / \partial x^r_l}{\partial u^r / \partial x^r_L}$ for $l = 1, \dots, L-1$. Then for $r \neq r'$ we assume $\partial \mathrm{MRS}^r(x^r, x^{r'}) / \partial x^{r'} = 0_{L-1,L-1}$ or in other words the marginal rates of substitution for own good consumption does not depend on the spouse's consumption. A sufficient condition for this would be for example separability of the form $u^s(x^r, x^{r'}) = G(g(x^r), x^{r'})$ for any two differentiable, increasing, real-valued functions G and g. While this assumption allows for positive consumption externalities, it restricts the way behaviour of a person is modified when entering or exiting a relationship. For example it rules out non-cooperative strategic behaviour of individuals within a couple. However, this assumption is still compatible with popular specifications such as the Beckerian caring model with altruistic preferences [Becker, 1981], in which utilities of one spouse are defined in terms of own-good consumption and the utility of the spouse, i.e. $u^r(x^r, x^{r'}) = W(U^r(x^r), U^{r'}(x^{r'}))$ where U^f and U^m are real-valued subutility functions with the usual properties and W is a strictly increasing, differentiable real-valued function. In addition to this separability assumption, in order to learn from single individuals we have to make the assumption that preferences of individuals do not change with respect to household composition, i.e. if they transition from being single to being in a couple or vice versa. We call this a *stable preference assumption* (see Dunbar et al. [2013a] and Vermeulen et al. [2006] for a detailed discussion). In what follows we will discuss how we can test this assumption non-parametrically.

The following observation will prove to be very useful in our approach. In order to check revealed preferences it is not necessary to consider the continuous distribution of consumption choices supported on the respective budget set. Rather is it sufficient to discretize the choice space by splitting the budget sets into regions which contain sufficient information to check revealed preference statements. The benefits of such

[3]This adds to the general observation that revealed preference restrictions often do not have much power to reject optimizing behaviour of individuals [Beatty & Crawford, 2011].

a discrete choice set are two-fold. First, it allows us to construct hypothetical household types, even if we only observe households with different compositions, i.e. single households and households consisting of two spouses. Secondly, it naturally introduces a notion of unobserved heterogeneity in which stochastic preference parameters of infinite dimensionality, representing a heterogeneous household, are mapped to an element of a finite-dimensional discrete choice space characterizing choices within and across different price regimes.

We will now discuss why it is sufficient to conclude about rationality of the population while only observing marginal distributions of choices for households with different compositions. Let the choice spaces for single males, single females and couples be denoted by $\mathcal{X}^m = \mathcal{X}^f$ and $\mathcal{X}^c$, respectively[4]. Each element of the choice space includes a sequence of choices in at least three periods, which we require to test the collective model. Let the probability that option ξ_j is chosen within a household of a given household composition r be denoted as $\pi(\xi_j|\mathcal{X}^r)$. We call this a stochastic choice in situation $\mathcal{X}^r$ and will often refer to it as the marginal distribution of choices under a given household composition. In order to identify these conditional probabilities we require panel data in which we observe choices of singles of both genders and couples. In addition to this, we require a time-homogeneity assumption of preferences such that we can treat different periods as different price regimes. To be more precise we have to assume that preferences do not change over time, such that we can treat the heterogeneity of choices between periods $t \in I_T$ to be a consequence of facing different prices p_t rather than a change in preferences over time[5].

To fully characterize households, we consider the product space $\mathfrak{X} = \mathcal{X}^c \times \mathcal{X}^m \times \mathcal{X}^f$ representing all possible combinations of choices for different household compositions. Each element contains all relevant information to construct the three-tuple (R^c, R^m, R^f), due to our separability and preference stability assumption, which ensures that single male and single female households are informative for the respective spouses' behaviour within a couple. A natural way to view this fully characterized type space is

[4]We will discuss how these choice spaces are constructed and how we can encode observed choices under different household compositions in detail in the next section.

[5]Note that this assumption can likely be relaxed to continuous demands of the form $\tilde{x}_{it} = \tilde{x}(p_t, \varepsilon_i) + \varepsilon_{it}$, where ε_i is an element of a general probability space capturing unobserved heterogeneity and ε_{it} is a period specific idiosyncratic taste shock, when projecting observed demands onto budgets of periods t. Identification of such a specification is treated in Evdokimov [2010].

via a clustered graph[6] where the nodes represent discrete consumption decisions and the partitioning is such that there are three disjoint classes each representing the set of discrete decisions under a given household composition – single female, single male and couple. By construction each node within a class is then connected to exactly one node of every other class. A hypothetical household composition is than characterized by a path of nodes connecting all three classes. Each of these paths can be classified based on its compatibility with the collective axiom. We denote the exhaustive set of paths that are collectively rational by $\mathfrak{X}^0$ but will be more precise about the exact partitioning later this section. The links between the nodes are not directly observed from data and thus neither are the paths.

However, using the principle of stochastic preferences [McFadden & Richter, 1991] we can ask the question whether there exists a probability measure ν over these paths supported on the space of fully characterized rational household preference relations $\mathfrak{X}^0$ that rationalizes the observed stochastic choices $\pi(\xi_j|\mathcal{X}^r)$ for $r \in \{c, m, f\}$. With this construction the (heterogeneous) choice function $\mathcal{X} \mapsto \Xi(\mathcal{X})$ determining a decision rule for a given state of the world is a function of household composition $\mathcal{X} \in \{\mathcal{X}^c, \mathcal{X}^m, \mathcal{X}^f\}$ rather than one of price regimes as in Kitamura & Stoye [2013]. We say that the stochastic choice π is stochastically rational if for all household compositions $\mathcal{X}^r$ for $r \in \{c, m, f\}$ it holds that $\pi(\xi_j|\mathcal{X}^r) = \nu(\{\Xi \in \mathfrak{X}^0 : \xi_j = \Xi(\mathcal{X}^r)\})$. Intuitively if the choices in different states of the world can be rationalized by a probability distribution over a set of rational households, we can say that the population is rational with respect to the decision rule Ξ.

To fix ideas for the testing procedure we propose in this paper we will now discuss which decision rule Ξ is appropriate. For this, we partition the universe of types $\mathfrak{X} = \mathfrak{X}^{\text{collective}} \cup \mathfrak{X}^{\text{alternative}}$, where the set $\mathfrak{X}^{\text{collective}}$ contains all hypothetical household types for which there exists a feasible consumption allocation which is consistent with the collective axiom as in Cherchye et al. [2007]. By also taking the actual respective preference relations of both spouses for each of these hypothetical types into account we can check whether a type is actually consistent with the collective axiom and denote this subset as $\mathfrak{X}^0 = \mathfrak{X}^{\text{collective}} \setminus \mathfrak{X}^1$, where $\mathfrak{X}^1$ is the complement of $\mathfrak{X}^0$ and consists of the cases which satisfy the necessary conditions of collective rationality but are not consistent with the collective axiom if preference relations from singles are added. Using

[6] Technically a k-partite graph with $k = 3$ which is a graph whose vertices can be partitioned into k disjoint sets, with none of the elements within each set being adjacent.

this particular partition, one could think of three different tests. Firstly, one could test $\mathfrak{X}^{\text{collective}}$ against $\mathfrak{X}^{\text{alternative}}$. In fact, this is what Cherchye et al. [2007] do which does not require single data since such a test can be based on aggregate consumption only and the whole joint distribution of such choices is directly identified from data. Secondly, adding single data and assuming separable and stable preferences one could test $\mathfrak{X}^0$ against $\mathfrak{X}^1 \cup \mathfrak{X}^{\text{alternative}}$ to obtain a stronger test of the collective model compared to the previous one. While this test has more power, it comes with the drawback of only applying to a separable caring-type model. Finally, one could drop all cases that are not collectively rational and consider $\mathfrak{X}^0$ and $\mathfrak{X}^1$ to answer the question whether the stable preference assumption holds. The latter two questions can both be answered using the tools we develop in this paper. In the empirical section of this paper we will however only focus on the very last one. If we find that among the collectively rational paths it is not possible to rationalize observed choice probabilities using the types belonging to the set $\mathfrak{X}^0$, then we can conclude that the hypothesis of stable preferences does not hold, since both $\mathfrak{X}^0$ and $\mathfrak{X}^1$ are disjoint subsets of the collectively rational types satisfying $\mathfrak{X}^{\text{collective}} = \mathfrak{X}^0 \cup \mathfrak{X}^1$. Before we will discuss how our inference can be based on finding a probability measure ν which rationalizes the observed choices probabilities π in Section 2.4, we will have to define the type space and show how we can formalize the decision rules in the next section.

Encoding

In order to characterize types, note that the Collective Axiom restricts not only aggregate household consumption characterized by the vector $\tilde{x}_t^c$, but also individual consumption of each spouse, denoted by $\tilde{x}_t^f$ and $\tilde{x}_t^m$. We start by defining and encoding the choice space of a given continuous consumption outcome $\tilde{x}_t^r$ for $r \in \{c, m, f\}$ in a given period $t \in I_T$. For this, take a family X_t of J_t disjoint non-empty sets denoted by $x_{j|t}$ which form a partition of a given (normalized) budget set: $B_t = \bigcup_{j=1}^{J_t} x_{j|t}$. To characterize discrete demands we then consider the sequence $\left\{\tilde{x}_{i,t}^r \in x_{j|t}\right\}_{j \in I_{J_t}}$ instead of $\tilde{x}_{i,t}^r$ itself. To keep the dimension of this discrete choice space as small as possible, we should make sure that we pick the coarsest partition that contains all the required information to verify the collective axiom. It turns out that in order to answer revealed preference statements it is sufficient to recursively split a budget set B_t where it intersects with other budget sets $B_{t'}$ for all $t \neq t'$.

Example. Although the notation and treatment is general we will mostly focus on the simplest possible setting, which still allows us to distinguish the collective model from the unitary model, namely the one with three goods and three price regimes. Figure 2.1 shows an example of such a three-good economy with three price-regimes characterizing budgets $B_{red}, B_{blue}, B_{green}$.

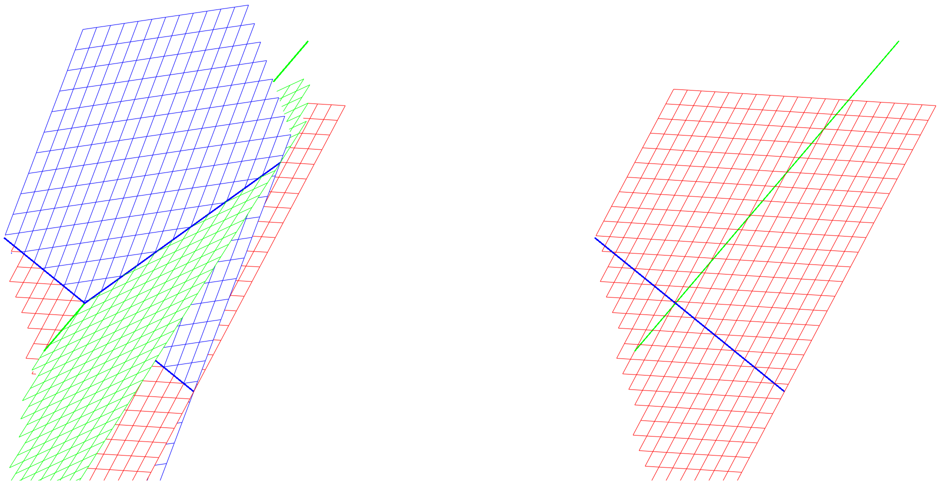

Figure 2.1: Three intersecting budget sets $B_{red}, B_{blue}, B_{green}$ with three goods

If all budgets intersect as in this example we get $J_t = 4$ possible discrete consumption choices for each budget set B_t where $t \in \{\text{blue, red, green}\} = I_T$. In the figure on the right hand side the green and blue budgets are removed and only the lines in which they intersect with the remaining red budget are plotted, showing that there are four quadrants which we will $x_{NW|red}, x_{NE|red}, x_{SE|red}$ and $x_{SW|red}$. We will repeatedly come back to this example setting to demonstrate our approach. □

For each budget set $t \in I_T$, following Kitamura & Stoye [2013] we encode the consumption choice indexed by $j \in J_t$, which is geometrically represented by a polyhedron, or in other words by an I_T-dimensional vector of inequalities

$$x_{j|t} = \begin{cases} +1 & \text{if } p_t\tilde{x}_j \geqslant 1 \\ 0 & \text{if } p_t\tilde{x}_j = 1 \\ -1 & \text{if } p_t\tilde{x}_j \leqslant 1 \end{cases} \qquad \text{for } t \in I_T$$

and denote the choice $x_{j|t}$ by *patch* j in period t. Let the operator which encodes a given $\tilde{x}$ as an element of X, the space of $(L-1)$-dimensional polyhedra, be denoted as $\mathfrak{D} : \mathbb{R}^{L-1} \to X$, where $\mathfrak{D}$ stands for *discretization*.

Example (cont.)**.** In our example the encoding of the choice set for the red budget, shown on the right hand side of Figure 2.1 is

$$X_{\text{red}} = \{x_{\text{NW}\,|\,\text{red}}, x_{\text{NE}\,|\,\text{red}}, x_{\text{SE}\,|\,\text{red}}, x_{\text{SW}\,|\,\text{red}}\}$$

$$= \left\{x_{\text{red, below blue, above green}}, x_{\text{red, below blue, below green}}, x_{\text{red, above blue, above green}}, x_{\text{red, above blue, below green}}\right\}$$

$$= \{(0,-1,+1),(0,-1,-1),(0,+1,+1),(0,+1,-1)\}.$$

The choice sets X_{blue} and X_{green} can be defined in a similar matter. □

We will now extend this choice characterization of Kitamura & Stoye [2013] and denote a choice under a given household composition as a *choice path* which is defined as the Cartesian product over choices per budget B_t for $t \in I_T$. We start with single households which will later be used to represent a member within a household. An element of the set of choice paths for $\mathcal{X}^r$, which fully characterizes the preference relations R^r for $r \in \{m, f\}$, can be defined as:

$$\xi_i^r = (\mathfrak{D}\,(\tilde{x}_{i,t}^r))_{t\in I_T} \in \mathcal{X}^r := \prod_{t\in I_T} X_t = \bigcup_{j=1}^{\prod_{t\in I_T} J_t} \xi_{j|r}.$$

Note again that in order to characterize this choice we need panel data and a time homogeneity assumption of preferences.

For aggregate demands, which we will only observe in couples, it can easily be seen that if we only discretized demands for a given period, some relevant information, namely on the sum of pairs of demands to make statements about items (iv) and (v) of the axiom, would be lost. In order to check these two parts of the axiom we therefore additionally have to store with each decision $\mathfrak{D}\,(\tilde{x}_{i,t}^c)$ at time $t \neq t' \neq t''$ the discretizations $\mathfrak{D}\,(\tilde{x}_{t'} + \tilde{x}_{t''})$. Note that due to the fact that the discretization operator we defined is not linear, we cannot simply add up consumptions of two periods since the quantities are encodings of discrete choices rather than continuous quantities and thus $\mathfrak{D}\,(\tilde{x}_t^c + \tilde{x}_{t'}^c) \neq \mathfrak{D}\tilde{x}_t^c + \mathfrak{D}\tilde{x}_{t'}^c$. Thus, we will treat them as additional discretized consumption outcomes. In a three period setting, this would be one additional encoded consumption vector, the encoding of which is treated in the example below. For four periods, we would have to compare consumption of period t with the sum of period t' and t'', the sum of period t'' and t''', as well as the sum of periods t' and t'''. Thus we can define the choice set for a couple $\mathcal{X}^c$ using the following discrete representation

$$\xi_i^c = (\mathfrak{D}(\tilde{x}_{i,t}^c), \mathfrak{D}(\tilde{x}_{i,t'}^c + \tilde{x}_{i,t''}^c))_{t,t',t''\in I_T, t\neq t', t'\neq t''} \in \mathcal{X}^c = \prod_{t\in I_T} X_t \times \mathcal{X}_t = \bigcup_{j=1}^{\prod_{t\in I_{T+\kappa}} J_t} \xi_{j|c}$$

where $\mathcal{X}_t$ represents the space of all $\kappa = \frac{1}{2}(T-1)(T-2)$ encodings of double-sums, which for the sake of the exposition will not be made explicit but will become apparent for the case of three goods and three budgets in the following example.

In the final step, the question that remains to be addressed is whether and how we can construct a household choice space that includes aggregate household consumption choice as well as individual male and female consumption. In other words we want to construct hypothetical types based on the sets of choices per household composition which fully characterizes couple preference relations and which we can then check against the conditions the collective axiom imposes on them. Such a hypothetical type space can be characterized by the Cartesian product $\mathfrak{X} = \prod_{r\in\{c,m,f\}} \mathcal{X}^r$.

Example (cont.). In our three good, three budget economy an example of a path would look as follows, containing six choices for the couple and three choices for the respective singles

$$\Xi = \begin{pmatrix} \xi^c \\ \xi^m \\ \xi^f \end{pmatrix} = \left\{ \begin{array}{c} \big(x_{\text{red, below blue, above green}}, x_{\text{above red, blue, above green}}, x_{\text{below red, above blue, green}}, \\ x_{\text{red, above blue+green}}, x_{\text{blue, below red+green}}, x_{\text{green, above red+blue}}\big), \\ \big(x_{\text{red, below blue, above green}}, x_{\text{above red, blue, below green}}, x_{\text{above red, above blue, green}}\big), \\ \big(x_{\text{red, above blue, below green}}, x_{\text{below red, blue, above green}}, x_{\text{above red, above blue, green}}\big) \end{array} \right\}$$

$$= \left\{ \begin{array}{llll} (& (0,-1,+1), & (+1,0,+1), & (-1,+1,0), \\ & (1,\ 0,\ 0), & (0,-1,\ 0) & (0,\ 0,+1)) \\ (& (0,-1,+1), & (+1,\ 0,-1), & (+1,+1,\ 0)), \\ (& (0,+1,-1), & (-1,\ 0,+1), & (+1,+1,\ 0) \quad) \end{array} \right\}$$

Example. Using the definition of the discretization operator we can go back to the set of inequalities characterizing revealed preference relations of a hypothetical household:

$$\begin{array}{lll|ll|ll} p_b x_r \leqslant 1, & p_g x_r \geqslant 1, & p_r(x_b + x_g) \geqslant 1 & p_b x_r^m \leqslant 1, & p_g x_r^m \geqslant 1, & p_b x_r^f \geqslant 1, & p_g x_r^f \leqslant 1 \\ p_r x_b \geqslant 1, & p_g x_b \geqslant 1, & p_b(x_r + x_g) \leqslant 1 & p_r x_b^m \geqslant 1, & p_g x_b^m \leqslant 1, & p_r x_b^f \leqslant 1, & p_g x_b^f \geqslant 1 \\ p_r x_g \leqslant 1, & p_b x_g \geqslant 1, & p_g(x_r + x_b) \geqslant 1 & p_r x_g^m \geqslant 1, & p_b x_g^m \geqslant 1, & p_r x_g^f \geqslant 1, & p_b x_g^f \geqslant 1 \end{array} .$$

This set of inequalities contains all the necessary information to check whether or not this choice is consistent with the Collective Axiom of Revealed Preference. □

Using the necessary conditions from the Collective Axiom we can now readily check the consistency of each type $\Xi \in \mathfrak{X}$. Since we use necessary conditions for collective rationality, we consider a hypothetical household composition Ξ to be not consistent with the collective model and the stable preference assumption if $\Xi \notin \mathfrak{X}^0$ and not consistent with the collective model at all if $\Xi \notin \mathfrak{X}^0 \cup \mathfrak{X}^1$. This can be checked using the encoded inequalities for hypothetical household type. Once we have classified these paths we obtain every possible type in the heterogeneous population of households.

Computation and Inference

In Section 2.3 we discussed the characterization of choices and constructed the type space. In what follows, we will discuss how to check rationality of each type and how to relate this type characterization to observed choice probabilities in each period and construct a test statistic, which allows us to test the stable preference assumption or to empirically distinguish the collective model under the stable preference assumption from a non-collective alternative or an irrational population.

Since with increasing periods, the complexity of checking the axiom increases exponentially (order $\mathcal{O}(2^N)$) due to the necessity of pair-wise comparison of all periods we implement a tree crawling algorithm that takes all combinations of choices in an expanding window of periods (levels). Intuitively if after checking the e.g. first two periods we already find a path to be irrational, there is no need to check all the child branches of the subtree. For a given level of the tree, we construct a graph representing preference relations which we immediately get from our encoded patches for which we have $(x_s^r p_t = -1 \iff x_s^r p_t \leqslant x_t^r p_t = 1) \Rightarrow x_t R_0^r x_s$. Given this preference relation R_0^r for each $r = \{c, m, f\}$, we can construct transitive closures R^f and R^m by applying the Warshall algorithm, of which our Haskell implementation can be found in Appendix 2.A.

Example (cont.). Graphs for the all required preference relations

$$R_0 = \begin{bmatrix} 0 & 1 & 0 \\ 0 & 0 & 0 \\ 1 & 0 & 0 \end{bmatrix} \qquad R_0^m = \begin{bmatrix} 0 & 1 & 0 \\ 0 & 0 & 1 \\ 0 & 0 & 0 \end{bmatrix} \qquad R_0^f = \begin{bmatrix} 0 & 0 & 1 \\ 1 & 0 & 0 \\ 0 & 0 & 0 \end{bmatrix}$$

with transitive closures

$$R^m = \begin{bmatrix} 0 & 1 & 1 \\ 0 & 0 & 1 \\ 0 & 0 & 0 \end{bmatrix} \qquad R^f = \begin{bmatrix} 0 & 0 & 1 \\ 1 & 0 & 0 \\ 0 & 0 & 0 \end{bmatrix}.$$

In this example we are dealing with a preference relation R_0^m of a person who prefers good one over good two and good two over good three. Thus he must also prefer good one over good three, for which there is not contradicting revelation of preferences. Thus this person is rational. The preference relation of R_0^f also represents a rational person, who prefers good one over good three and good two over good one.

Note that this implies that both individuals prefer good one over good three, but aggregate household consumption represented by R_0^c revealed that that the household chose good three over good one; a violation of the collective axiom of revealed preferences. □

Using a graph representation of the preference relations, the CARP can be checked by directly applying standard boolean logic on the elements of the preference relation. Once we have classified each possible type as rational under the stable preference assumption or rational using only the necessary conditions imposed on aggregate consumption, or not collectively rational at all we can encode every household of the heterogeneous population as one of the $|\mathfrak{X}|$ types.

Example. In our three-good economy we have $|\mathfrak{X}^0| = 2{,}996$ fully characterized household types who are collectively rational under the stable preference assumption. $|\mathfrak{X}^1| = 475{,}136$ are consistent with the collective axiom based on the necessary conditions using only aggregate household consumption data. In total we have $|\mathfrak{X}| = 2{,}097{,}152$ types. This leaves us with about 22.7% collectively rational types. From this we should not necessarily conclude a restrictive nature of the collective model, since for a given range of budget planes only a subset of the total choice set would actually be feasible (e.g. have positive demands).

In order to link our hypothetical types to actual observed data we use the result of McFadden & Richter [1991]; McFadden [2005] who show that the question of rationality of a heterogeneous population boils down to whether or not we can find a convex combination of rational choice types characterized by columns of a matrix A that rationalizes observed choice probabilities π in the following way:

Definition 2.2. Collectively rational heterogeneous population

A population is rational if there exists a vector $\nu \in \Delta^{|\mathfrak{X}^0|}$ which satisfies $A\nu = \pi$ where Δ^M is the M-dimensional probability simplex.

In order to construct such a matrix representation A, based on which we can conduct statistical inference, we consider a matrix with $\sum_{r\in\{c,m,f\}} |\mathcal{X}^r|$ rows and $|\mathfrak{X}^0|$ columns, where $|\mathcal{X}^r|$ is the number of choices a household can make under a given composition. Now we split all columns $A_{\blacksquare,m}$ with $m \in I_{|\mathfrak{X}^0|}$ of the matrix A into 3 blocks of respective length $|\mathcal{X}^c|$, $|\mathcal{X}^m|$ and $|\mathcal{X}^f|$ and denote each block by $A_{r,\blacksquare,m}$. If household type $m \in I_{|\mathfrak{X}^0|}$

picks option j under composition $r \in \{c, m, f\}$ then $A_{r,j,m} = 1$ and zero otherwise. In the graph interpretation of the type space which was discussed in Section 2.2, a block represents a class or cluster in the graph. For a given class and a given path where the latter is represented by a column of the matrix A, a row value of 1 indicates the active node within that path.

The second ingredient is a vector of observed choice probabilities which we observe by mapping observed continuous consumption onto the defined type space for each household $i \in I_N$. Thus we can define the vector π with $\sum_{r\in\{c,m,f\}} |\mathcal{X}^r|$ rows in which we collect observed choice probabilities. Partitioning π the same way as a column A_m, we define $\pi_{r,j} = \frac{1}{N}\sum_{i=1}^{N} \sum_{\xi\in\mathcal{X}^r} \mathbb{1}\{\xi_i = \xi\}$ where $\mathcal{X}^r$ is the households' choice space under composition $r \in \{c, m, f\}$ as defined above and ξ_i is the encoded observed choice of household $i \in I_N$, which can be either single or couple.

Example (cont.)**.** In our example the cardinality of $\mathcal{X}^r$ is $|\mathcal{X}^r| = \prod_{t\in I_T} J_t = 4 * 4 * 4 = 64$ for $r \in \{m, f\}$ representing single males and single females, respectively. For couples we have to store double-sums for which we have $J_{T+\kappa} = 2^3$ different possibilities with $\kappa = \frac{1}{2}(T-1)(T-2) = 1$ which results in $\prod_{t\in I_{T+\kappa}} J_t = 4{*}4{*}4{*}8 = 512$ choices. Hence we obtain a matrix A with $512 + 64 + 64 = 640$ rows and 2,996 columns representing rational types under the stable preference assumption. The vector π is a vector of choice probabilities of the population of the same dimension: 640. While this might seem high-dimensional we note that this matrix is very sparse. In fact it only has $3|\mathfrak{X}^0| = 3{*}2{,}996$ non-zero items.

The way the matrix A is constructed ν is not point-identified since A is far from full column rank since $|\mathfrak{X}^0| \gg \sum_{r\in\{c,m,f\}} |\mathcal{X}^r|$. However, we can base our test upon distances from the observed choice probabilities to the choice set representing the null hypothesis using McFadden & Richter [1991]; McFadden [2005] who provide a very useful equivalent formulation of the problem, namely:

$$\exists \nu \in \Delta^{|\mathfrak{X}^0|} \text{ s.t. } A\nu = \pi \iff \mathcal{J}_N := N \min_{\eta\in\{A\nu|\nu\geqslant 0\}} (\pi-\eta)^T \Omega (\pi-\eta) = 0$$

where Ω is a square weighting matrix. This is a quadratic problem with first-order cone constraint and its solution will be zero if and only if a ν rationalizes observed choice probabilities of a heterogeneous population. To see why $\nu \geqslant 0$ is sufficient for it to be on the probability simplex, note that for any solution of the quadratic problem we have $\eta = \pi$ and since $J = \iota^T\pi = \iota^T A\nu = J\iota^T\nu$ by construction, we get $\iota^T\nu = 1$. The factor N on the right hand side is a normalization used to construct a test-statistic based on the solution $\mathcal{J}_N$, which is unique and thus identified. In other words, we project observed choice probabilities π onto the linear cone represented by the constraint $\mathcal{C} = \{A\nu : \nu \geqslant 0\}$, and use the distance given by the projection residuals as a basis for our statistical inference in which we obtain the critical value based on a bootstrap. As such, the projection residuals are calculated very often and it will prove useful to rewrite $\mathcal{J}_N$ as the solution of a non-negative least squares problem and implement a fast algorithm for solving it.

It is easy to show that $\mathcal{J}_N$ can be expressed as the solution of the non-negative least squares problem (NNLS) such that

$$\min_{\eta \in \{A\nu | \nu \geqslant \underline{\nu}\}} N(\pi - \eta)^T \Omega (\pi - \eta) = 0 \iff \min_{\nu \geqslant 0} N \left\| \tilde{A}\nu - \tilde{b} \right\|^2$$

where $\tilde{A} = L^T A$, $\tilde{b} = L^T(\pi - A\underline{\nu})$ and L is a lower diagonal matrix from the Cholesky decomposition $\Omega = LL^T$.

There exist a range of approaches to solve such a program, including the widely-used[7] Lawson & Hanson [1995] algorithm, Landwebers gradient descent method [Johansson et al., 2006] and sequential coordinate-wise optimization [Franc et al., 2005]. The latter two are more efficient than the general sequential quadratic programming approach, requiring only $\mathcal{O}(k)$ computations instead of $\mathcal{O}(k^3)$, where $k = |\mathcal{X}^0|$ is the number of variables (rational types) in the NNLS problem. Due to the high dimensionality of the problem we use the Landweber method which we implement manually, in order to leverage the sparsity of the matrix A.

Since ν is not point-identified we have to follow a partial identification approach for inference. We will use the τ_N-tightened bootstrap estimator proposed by Kitamura & Stoye [2013]. Tightening is required since many of the inequality constraints describing the cone will be binding and it is well known that the bootstrap is not valid if the parameter is on the boundary of the parameter space [Andrews, 2000]. The estimation procedure can be summarized as follows. First we estimate η by projecting observed choice probabilities onto a τ_N-tightened linear cone constraint by minimizing

$$\widehat{\eta}_{\tau_N} = \arg \min_{\eta \in \{A\nu | \nu \geqslant \iota\tau_N\}} (\widehat{\pi} - \eta)^T \Omega (\widehat{\pi} - \eta)$$

and denoting the value of the quadratic function of at $\widehat{\eta}_{\tau_N}$ as $\widehat{\mathcal{J}_N^{\tau_N}}$.

The solution is then used to calculate for each bootstrapped $\widehat{\pi}^b$, where $b \in I_B$ and B is the number of bootstrap repetitions, the centered choice probabilities $\widehat{\pi}^b_{\tau_N} = \widehat{\pi}^b -$

[7] This algorithm is used by Matlab (`lsqnonneg`) and SciPy (`optimize.nnls`).

$\hat{\pi} - \hat{\eta}_{\tau_N}$ which we then use to approximate the empirical distribution $F_{\mathcal{J}_N}$ of $\mathcal{J}_N^{\tau_N}$ by optimizing for each $b \in I_B$

$$\mathcal{J}_{N,b}^{\tau_N} = \min_{\eta \in \{A\nu | \nu \geqslant \iota\tau_N\}} (\hat{\pi}_{\tau_N}^b - \eta)^T \Omega (\hat{\pi}_{\tau_N}^b - \eta).$$

Then under some weak regularity conditions [Kitamura & Stoye, 2013] the bootstrap is valid and we have for $\alpha \in (0, \frac{1}{2})$ and $\tau_N \sqrt{N} \to \infty$

$$\lim_{N\to\infty} \inf_{\pi \in \mathcal{C}} \inf \mathbf{P}\left(\mathcal{J}_N \leqslant \widehat{F_{\mathcal{J}_N}^{-1}}(1-\alpha)\right) = 1 - \alpha. \tag{2.1}$$

It should be noted, that we chose a slightly different tightening sequence than the one in the original paper. To be precise, we set $\tau_N = \frac{1}{H}\sqrt{\frac{\log N}{\underline{N}}}$ where $\underline{N} = N_f \wedge N_m \wedge N_c$ is the minimum number of available observations per household composition N_r for $r \in \{c, m, f\}$ in the sample. This differs from the original proposition by the factor $\frac{1}{H}$. To see why one might otherwise run into problems for finite samples, note that τ_N is constrained to satisfy $\sqrt{N}\tau_N \to \infty$. In addition to this if $\nu \geqslant \iota\tau_N$, by construction we have $J = \iota^T A\nu \geqslant \iota^T A\tau_N = J\iota^T\iota\tau_N = JH\tau_N$. Thus, combining these two conditions we get $\frac{1}{H} \geqslant \tau_N > \frac{1}{\sqrt{N}}$, which is not always feasible for a given sample size and number of types (N, H) if $\tau_N = \sqrt{\frac{\log N}{\underline{N}}}$ but holds by construction if we divide it by H.

Simulations

In this section we investigate the properties of our proposed test in a simulation setting. In particular we are interested in how much power it has to detect a violation of the stable preference assumption and whether or not it has a reasonable frequency of false positives.

Since specifying a parametric continuous demand system requires at least five goods to impose the SNR(S-1) condition on the Slutsky matrix and distinguish the collective model from the unitary model, we will not sample continuous demands as functions of prices and individual budget constraints, but rather draw our sample directly from the discrete choice space[8]. This should be interpreted as a continuous uniform distribution of choices on different budget planes, where the relative prices are such that the partitions of the budget planes are of equal size. Recall that we test this against the set of households which are consistent with the necessary conditions of the collective axioms based on aggregate consumption but not consistent when single data

[8]A revealed preference based setting allows us to test the restrictions of the model with only three goods [Cherchye et al., 2007], whereas Browning & Chiappori [1998] need five goods.

and the stable preference assumption is added. This set is denoted by $\mathfrak{X}^1$ and we have $\mathfrak{X}^{\text{collective}} = \mathfrak{X}^0 \cup \mathfrak{X}^1$. If we reject the null hypothesis that both the collective axiom and the stable preference assumption holds, by excluding all irrational paths $\mathfrak{X} \setminus \mathfrak{X}^{\text{collective}}$, we must conclude that the stable preference assumption does not hold. To control the proportion of households for whom this is the case (our data generating process) we introduce the parameter p which specifies the probability[9] that a particular choice is both collectively rational and satisfies the stable preference assumption $p := \mathbf{P}(x \in \mathfrak{X}^0)$. By only considering collectively rational choices in our simulations we thus have $1 - p = \mathbf{P}(x \notin \mathfrak{X}^0) = \mathbf{P}(x \in \mathfrak{X}^1)$ by construction.

Our simulation setting is as follows. We consider $S = 100$ samples of size $\underline{N} \in \{500, 1000, 2000\}$ where $\underline{N} = N_f = N_m = N_c$ such that $N = 3\underline{N}$ in a minimal setting of $T = 3$ periods which we construct by drawing $\lfloor \underline{N}p \rfloor$ indices from the space of collectively rational choice paths $\mathfrak{X}^0$ for which the stable preference assumption holds and $\lceil \underline{N}(1-p) \rceil$ indices from the space of collectively rational types $\mathfrak{X}^1$ which does not satisfy the assumption. Based on a sample of choice paths, we then calculate the choice probabilities $\widehat{\pi}$ accordingly. For estimation we only use the marginal distribution of choices of each sample of household compositions and draw $B = 100$ samples from the respective empirical distributions (i.e. with replacement) to calculate $\pi^b_{\tau_N}$ and estimate the empirical distribution of the test statistic $\mathcal{J}^{\tau_N}_{N,b}$. These simulations are repeated for $p \in \{0.75, 0.85, 0.9, 0.95, 0.975, 0.99, 1.00\}$.

Figure 2.2 shows the power of our test against the the non-stable preference alternative as a function of p, with sample-size $\underline{N} = 500$ for the left hand side graph, and $\underline{N} = 1000$ for the right hand side graph, respectively. We use monotone cubic splines to interpolate between the actual simulation results, which are marked as solid dots. To be more precise, the respective functions refer to sample rejection frequencies using the rejection rule $J \mapsto \mathbb{1}\left\{J > \widehat{F^{-1}_{\mathcal{J}_N}}(1-\alpha)\right\}$ for $\alpha \in \{0.01, 0.05, 0.10\}$. In addition to this, we also observe that as $\underline{N}$ increases the power of our test improves and is able to correctly reject the hypothesis of a collectively rational population already at small proportions p.

The intercepts of these functions should be interpreted as the proportion of false positives (type I errors), since they correspond to the case where everyone is ratio-

[9] This rationality parameter is similar as for example λ in Dette et al. [2016] which specifies the population's deviation from Slutsky symmetry.

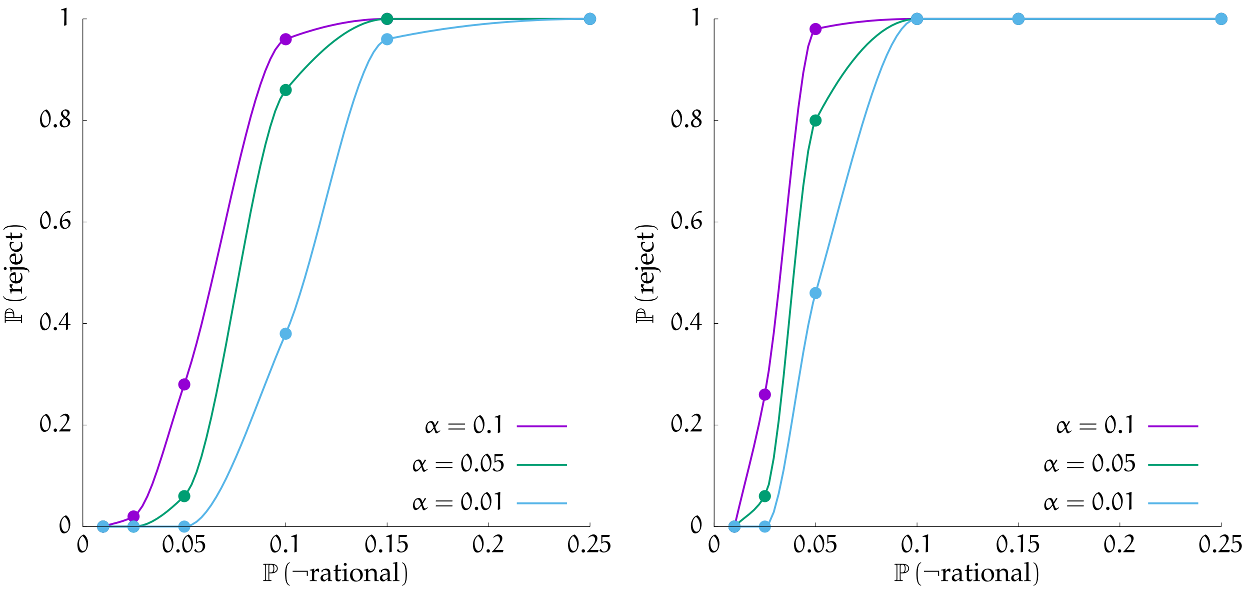

Figure 2.2: Power function for $N = 1{,}500$ (l.h.s) and $N = 3{,}000$ (r.h.s.)

nal. One might expect that for a correctly sized test the empirical rejection frequencies should tend to α. However, given our partial identification procedure we have a composite null hypothesis, *i.e.* the probability of a type I error should be at most α as defined in equation (2.1). To see this note that every vector of "true" choice frequencies denoted by π_0 that is in the interior of the cone will have projection residuals of length zero. Bootstrapping out of $\widehat{\pi}$ which tends to π_0 using the usual regularity properties could then lead to a confidence interval which is always entirely in the interior of the cone and we would never wrongly reject the null hypothesis. This also implies that in such a case our bootstrap distribution is degenerate and has mass one at point zero.

In our Monte Carlo setting in the case where $p = 1.0$ we randomly select types from the type-space $\mathfrak{X}^0$, satisfying collective rationality. Thus the "true" parameter vector ν_0 is assumed to have a uniform distribution over the probability simplex and the worst case – namely to get a ν such that $\pi_0 = A\nu$ is on the boundary of the cone with respect to any of its dimensions – occurs with measure zero.

Thus, in order to evaluate whether the size of our test is correct under the test's minimax strategy, we have to construct a worst case. For this, note that the test is constructed in a way that considers hypothetical types by taking combinations of possible household choice behaviour per price regime over a range of price regimes. To fix notation, we will call two collectively rational choice paths *similar* if there is at least one element in the product space spanned by these two paths which is an element of the space of collectively rational paths that do not satisfy the stable preference hypothesis. We will then construct worst cases by specifying a distribution over n_0 such similar paths. To make sure that our π_0 is on the boundary of the cone in all dimensions, *i.e.* on the cusp, we shift the cone by manually controlling the tightening parameter τ_N according to this distribution. Figure 2.3 shows simulation results for two such worst case scenarios with 5 similar paths and 2 similar paths, respectively.

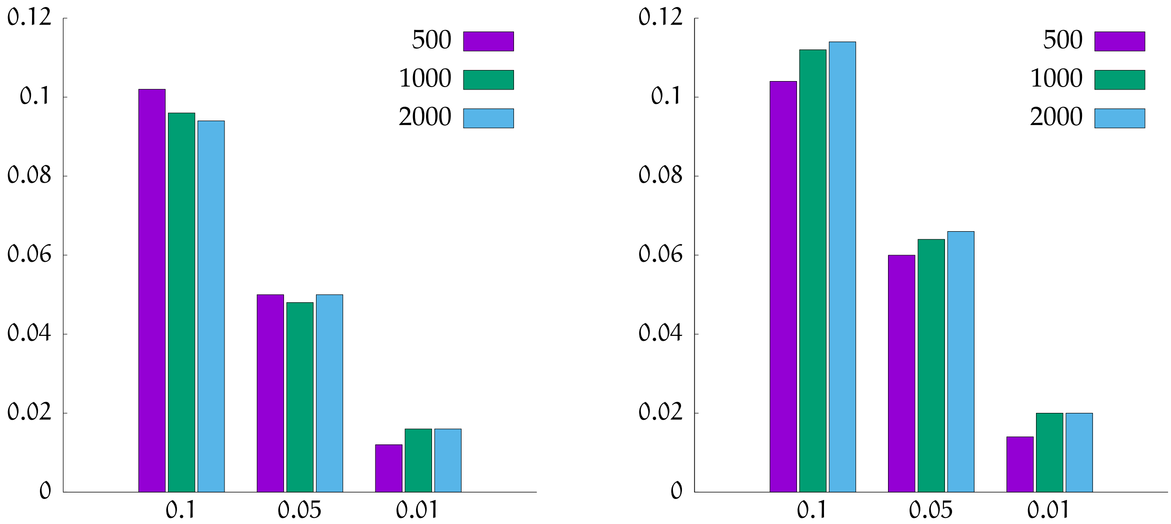

Figure 2.3: Type I error for $n_0 = 5$ (l.h.s) and $n_0 = 2$ (r.h.s.) worst-case paths

While both are asymptotically valid from a theoretical point of view, it is not surprising that for finite samples the size for the case with a larger number of worst case paths behaves worse than the case in which there are fewer worst case paths. Since the properties of the test are based on an asymptotic argument, we should see the empirical frequency of false positives tending to the respective α which define the rejection rules and are plotted on the x-axis. The results are what one would expect, with all sample sizes being reasonably accurate. Since in a well-behaved test false-positives are by definition rather rare events, in order to minimize simulation uncertainty, we increased the number of Monte Carlo repetitions to $S = 500$ and the number of bootstrap repetitions to $B = 200$, which greatly increased computational complexity due to the high dimensionality of the testing problem.

Conclusion

In this paper we use tools from both the discrete non-parametric collective consumption literature and the one from stochastic random utility modeling to construct a test of the stable preference assumption which is often used in collective model, stating that consumption preferences do not change as individuals transition between relationship states. For this we constructed hypothetical household types that satisfy revealed preferences restrictions the collective model imposes on observed household demands and showed how we can exploit information from single households in order to non-parametrically test the stable preference hypothesis. We provided simulation evidence that our test has power against the alternative hypothesis of non-stable preferences. In addition to this, we discussed worst cases and showed that the size of the test is correct under such scenarios. A natural extension of this paper would be to impose the single preference assumption and use it to identify primitives of the collective model such as the sharing rule. In addition to this, we think that the assumption of time-stability of preferences can be relaxed in this context when projecting observed continuous demands of households of different compositions onto the common budget planes using panel data, by allowing for additive idiosyncratic and period-specific errors in the structural demand functions for which general identification results exist in the literature [Evdokimov, 2010].

Algorithms

```
type PrefRel = [[Bool]]

warshall :: Int -> Int -> PrefRel -> Bool
warshall start end graph = adjacent (start, end, length graph) graph

adjacent :: (Int, Int, Int) -> PrefRel -> Bool
adjacent (i, j, 0) g = g !! (i-1) !! (j-1)
adjacent (i, j, k) g = adjacent (i, j, k-1) g || (adjacent (i, k, k-1) g && adjacent (k, j, k-1) g)

transitiveClosure :: PrefRel -> PrefRel
transitiveClosure adj = toLists $ uncurry matrix dim (\ (i, j) -> warshall i j adj)
     where dim = (length adj, head (map length adj))
```

Algorithm 2.1: Warshall Algorithm

CHAPTER 3

Smooth Transition GARCH Models

Introduction

With increasing regulatory efforts and new standards for determining capital requirements for financial institutions, methods for estimating conditional quantiles and conditional volatilities have been getting significantly more attention. While a vast amount of models for conditional variance has been developed with Engle [1982] and Bollerslev [1986] leading the way, only very few models exist for directly estimating conditional quantiles, the main ones being the conditional quantile ARCH model [Koenker & Zhao, 1996] and the Conditional Autogregressive Value at Risk (CAViaR) model by Engle & Manganelli [2004] which can be interpreted as the conditional quantile analogue of the GARCH model. For a comprehensive discussion of different value at risk estimators and their respective merits see Xiao et al. [2015]. Although there is a direct link between the two approaches, calculating the τ^{th} conditional quantile from a conditional variance estimate usually requires making a distributional assumption on the error terms. A wrong choice of distribution can influence the estimates and their interpretation to a large extent, which is particularly harmful for tail estimation such as the *Value at Risk*. On the other hand, it is often seen in financial time series that dynamics with respect to positive and negative news are different, which can be theoretically justified by the leverage effect and volatility feedbacks [Andersen & Bollerslev, 2006] or behavioural factors such as loss aversion [McQueen & Vorkink, 2004]. In addition to this, time series may be subject to cyclical behaviour or time-dependent frequencies which are not captured by a linear model [Tong & Lim, 1980]. Thus it is beneficial and will improve the accuracy of our forecasts if we allow for such asymmetric dynamic behaviour. A very general approach to model asymmetric responses to past shocks is the smooth transition approach of Terasvirta [1992], which includes a threshold model [Tong & Lim, 1980] as a limit case.

In this paper, we introduce a smooth transition general autoregressive conditional quantile model in which we allow conditional quantiles to follow an autoregressive process that also depends on past conditional volatilities as in Engle & Manganelli [2004]. We allow for asymmetric responses by specifying two regimes, each represented by its own parameter vector. The active regime is determined by a transition function characterized by the location and scale parameters and a transition variable that can be both a lag of the dependent variable or an exogenous variable. Our paper is related to Xiao & Koenker [2009] who provide a method to estimate the CAViaR model without regime-switching by employing a three-stage procedure, first estimating an ARCH approximation of the model, followed by a minimum-distance estimation step to calculate conditional volatilities, which are then in a final step used for the estimation of the CAViaR model's parameters. The model and estimation procedure we propose can be seen as an extension of this to a regime switching-framework. In addition to this, we deviate from the original one as we merge the author's first and second steps using composite quantile regression [Zou & Yuan, 2008], which allows us to eliminate the second step by directly estimating global parameters defining conditional volatilities. Conditional upon the latter, we can then estimate the CAViaR parameters by using standard quantile regression techniques as in Koenker & Bassett [1978].

The idea of regime switching models in its most general form is well established in the context of conditional variance estimation, see Li & Li [1996], Gonzales-Rivera [1998], and Anderson et al. [1999] who use a self-exciting threshold, a smooth transition, and an asymmetric non-linear smooth transition specification, respectively. While some empirical research has been done on the topic of modelling regime-switching conditional quantiles, such as White et al. [2008] and Huang et al. [2009] who allow for asymmetric responses of autoregressive conditional quantiles, but who do not provide any theory, it seems that, compared to its conditional variance counterpart, the possibilities to model asymmetric responses of time series to positive and negative shocks are rather limited in the quantile regression framework. Although Engle & Manganelli [2004] propose an asymmetric version of the CAViaR model, namely a Glosten-

Jagannathan-Runkle (GJR) specification [Glosten et al., 1993], this only accounts for the case where the regime-switch is a threshold located at zero and also disregards any asymmetric impacts of past conditional quantiles, for which e.g. Nam et al. [2001] have found empirical evidence in economic and financial time series. An extension of this threshold model which also allows for two regimes with respect to past conditional quantiles is Gerlach et al. [2011], who use Bayesian estimation techniques to estimate the Value at Risk. Cai & Xu [2009] on the other hand take a somewhat different route and propose a nonparametric estimator which allows for a time-varying coefficient for Value at Risk estimation.

Compared to the threshold model a smooth-transition approach, as employed here, facilitates a higher degree of flexibility, allowing both the location and scale of the transition function to be model parameters. In addition to this, using a continuous transition function allows for an arbitrary combination of two regimes, which can also be interpreted as a range of regimes and can therefore not only account for different impacts below or above a certain changepoint, but also for different magnitudes around it. Further, compared to modelling conditional variances, quantile regression has the benefit of not requiring a distributional assumption about the innovations of the time series. While there exists a quantile regression estimation procedure for a quadratic form of conditional variance [Lee & Noh, 2013] we will assume a linear structure of conditional volatility instead [Taylor, 1986; Schwert, 1990]. This has proven to be less sensitive to outliers as shocks enter the conditional volatility in an absolute rather than in a squared form. It is well established that in GARCH models the latter leads to an over-prediction of future volatility levels [Klaassen, 2002]. Another benefit of such a linear specification is that, instead of requiring the existence of the 6^{th} moment in the case of a regime-switching model with a quadratic form, we only require the $(4+\delta)^{\text{th}}$ finite moment for our innovation distribution. A further advantage of quantile regression over conditional variance estimation results from the fact that maximum likelihood estimation of these highly non-linear models is avoided. For regime-switching conditional variance models, maximum likelihood estimation often faces serious convergence issues, especially if outliers are present, such that convergence is sensitive to the initial parameter value and the choice of the transition function. Some of these issues are investigated in Section 3.5. For a rigorous discussion thereof we refer the reader to Chan & McAleer [2003] and references therein.

Finally it should be clarified that the goal of this paper is to model asymmetries within time series, rather than structural breaks. These two approaches share a lot of similarities since they both allow for multiple parameter regimes. In structural break models however a transition from one regime to another usually occurs only once, whereas in self-exciting models the active parameter regime constantly switches depending on the transition variable. While technically the time index could be used as an exogenous transition variable in our approach and as such nest a break point model as a special case, it is not the purpose of this paper to detect structural breaks in time series.

This paper is structured into two main parts. The first part starts by specifying the regime-switching CAViaR model in Section 3.2, followed by a description of the proposed estimation procedure in Section 3.3 and its asymptotic properties in Section 3.4. This is followed by a computational part consisting of a comprehensive Monte Carlo study in Section 3.5 and an empirical application in Section 3.6 which demonstrates how the method can be used for empirical research.

Model Specification

Let u_t be a stochastic process defined on the real line, where the stationary sample process $\{u_t\}_{t=1}^{n}$ is observed and follows the standard conditional scale model

$$u_t = \sigma_t\left(z_t, \theta_0\right)\varepsilon_t, \tag{3.1}$$

where $\{\varepsilon\}_t$ are i.i.d. distributed with mean zero and finite variance according to a right-continuous distribution function $F_\varepsilon(x) = \mathbf{P}(\varepsilon_t \leqslant x)$ and $\sigma_t : \mathcal{F}_{t-1} \times \Theta_2 \longrightarrow \mathbb{R}_+$ with $\mathcal{F}_{t-1}$ denoting the σ-algebra generated by the process $\{u_s\}_{s=-\infty}^{t-1}$ up to time $t-1$ and Θ_2 as the parameter space. We will denote past observations up to $t-1$ by z_t, which are assumed to be independent of ε_t.

A specific structural assumption for $\sigma_t(z_t, \theta)$ is, for example, the standard quadratic GARCH(p,q) specification with $z_t = (\sigma_{t-1}^2, \ldots, \sigma_{t-p}^2, u_{t-1}^2, \ldots, u_{t-q}^2)$ and parameters $\theta = (\beta_0, \beta_1, \ldots, \beta_p, \gamma_1, \ldots, \gamma_q)$ such that

$$\sigma_t(z_t, \theta) = \left(\beta_0 + \sum_{i=1}^{p} \beta_i \sigma_{t-i}^2(z_t, \theta) + \sum_{j=1}^{q} \gamma_j u_{t-j}^2\right)^{1/2}. \tag{3.2}$$

In the conditional quantile model we propose, it will prove useful to specify the following absolute value alternative of the GARCH(p,q) model using

$$z_t = z_t^{pq} := (\sigma_{t-1}, \ldots, \sigma_{t-p}, |u_{t-1}|, \ldots, |u_{t-q}|)$$

and θ as defined above

$$\sigma_t(z_t, \theta) = \beta_0 + \sum_{i=1}^{p} \beta_i \sigma_{t-i}(z_t, \theta) + \sum_{j=1}^{q} \gamma_j |u_{t-j}|. \tag{3.3}$$

In order to introduce regime-dependency, it is assumed that the true conditional volatility process is generated according to the following general two-regime GARCH(p, q) specification

$$\sigma_t\left(z_t, \theta^{I}, \theta^{II}, \zeta\right) = G\left(\xi_t\left(z_t\right), \zeta, \eta\right) \sigma_t\left(z_t, \theta^{I}\right) + (1 - G\left(\xi_t\left(z_t\right), \zeta, \eta\right)) \sigma_t\left(z_t, \theta^{II}\right), \tag{3.4}$$

in which each regime is allowed to have different dynamics characterized by regime-specific parameter vectors θ^{I} and θ^{II}, respectively. The function $G : \mathbb{R}^2 \times \mathbb{R}_+ \longrightarrow \mathcal{R} \subseteq [0,1]$ represents the transition function that depends on the transition variable modeled as a known function of past observations $\xi : \mathcal{F}_{t-1} \longrightarrow \mathbb{R}$ and is parametrized by location parameter $\zeta \in \mathbb{R}$ and scale parameter $\eta \in \mathbb{R}_+$ determining the active regime. For notational convenience we will stack the transition parameters to the vector $\boldsymbol{\zeta} = [\zeta, \eta]$. We restrict the function ξ to be time-homogeneous and will refer to it as a transition variable $\xi_t := \xi(z_t)$ even though any scalar deterministic function of past data $z_t \in \mathcal{F}_{t-1}$ can be specified. One example, in case of daily data, would be the last week's average returns $\xi(z_t) = \frac{1}{5}\sum_{j=1}^{5} u_{t-j}$. In the standard quadratic asymmetric non-linear smooth transition GARCH model (which we will denote by ANST-GARCH) [Anderson et al., 1999] $\sigma_t(z_t, \theta^r)$ takes the form (3.2). For the model and estimation procedure we propose, we will however assume that the conditional volatility process for each regime follows the absolute value form defined in (3.3).

Assumption 3.1. The transition function must satisfy the following properties:

$$\lim_{\xi \to -\infty} G(\xi, \zeta, \eta) \to 0 \quad \text{and} \quad \lim_{\xi \to +\infty} G(\xi, \zeta, \eta) \to 1$$

is monotone, measurable and Lipschitz. Further $\partial^d G / \partial(\zeta, \eta)^d$ exists almost everywhere for $d = 1,2$, is bounded and Lipschitz with respect to ζ and η. In addition to this $\frac{\partial G}{\partial(\zeta, \eta)}$ is monotone or Lipschitz in ξ.

Standard choices for the transition function include:

(i) the logistic distribution function as used in Terasvirta [1992] for the Logistic Smooth Transition Autoregressive (LSTAR) model

$$G_{\text{logistic}} : \mathbb{R}^2 \times \mathbb{R}_+ \longrightarrow [0,1] : (\xi, \zeta, \eta) \longmapsto \left(1 + \exp(\eta^{-1}(\xi - \zeta))\right)^{-1},$$

(ii) the scale-invariant indicator function which reduces the model to the threshold version as in Li & Li [1996]

$$G_{\text{threshold}} : \mathbb{R}^2 \longrightarrow \{0,1\} : (\xi, \zeta) \longmapsto \mathbb{1}\{\xi \geqslant \zeta\},$$

(iii) and a bounded linear function with location centered between two cut-off points

$$G_{\text{linear}} : \mathbb{R}^2 \times \mathbb{R}_+ \longrightarrow [0,1] :$$
$$(\xi, \zeta, \eta) \longmapsto \left(\eta^{-1}\left(\xi - \zeta + \frac{1}{2}\eta\right)\right) \mathbb{1}\left\{\xi \in \left[\zeta - \frac{1}{2}\eta, \zeta + \frac{1}{2}\eta\right)\right\} + \mathbb{1}\left\{\xi \in [\zeta + \frac{1}{2}\eta, \infty)\right\}.$$

Our theoretical results are based on the class of transition functions defined by the restrictions in Assumption 3.1. While there is no doubt that other functions also satisfy Assumption 3.1, for the empirical part of the study we will restrict the set of transition functions to $\mathfrak{G} := \{G_{\text{logistic}}, G_{\text{linear}}\}$. In addition to this we will empirically evaluate what happens if the DGP follows the limit case with $G_{\text{threshold}}$. We will sometimes abbreviate the transition function as $G_t(\zeta) = G(\xi_t, \zeta, \eta)$.

Having defined the ANST-GARCH model, the shift to the quantile specification is straightforward. First, note that the τ^{th} conditional quantile of u_t is defined by

$$Q_{u_t}(\tau|\mathcal{F}_{t-1}) := \inf\left\{x \in \mathbb{R} : F_{u_t|\mathcal{F}_{t-1}}(x) \geqslant \tau\right\},$$

with $F_{u_t|\mathcal{F}_{t-1}} : \mathbb{R} \longrightarrow [0,1]$ being the conditional distribution function of u_t given all past observations such that $F_{u_t|\mathcal{F}_{t-1}}(x) = \mathbf{P}(u_t \leqslant x|\mathcal{F}_{t-1})$. It follows for the model defined in equation (3.1) that

$$\tau = \mathbf{P}(u_t \leqslant Q_{u_t}(\tau|\mathcal{F}_{t-1})) = \mathbf{P}(\sigma_t \varepsilon_t \leqslant Q_{u_t}(\tau|\mathcal{F}_{t-1})) = F_\varepsilon\left(\sigma_t^{-1} Q_{u_t}(\tau|\mathcal{F}_{t-1})\right)$$
$$Q_{u_t}(\tau|\mathcal{F}_{t-1}) = \sigma_t F_\varepsilon^{-1}(\tau). \tag{3.5}$$

Using this result and multiplying the ANST-GARCH model from equation (3.4) by $F_\varepsilon^{-1}(\tau)$, the final asymmetric non-linear smooth transition generalized autoregressive conditional quantile model (ANST-GACQ) is obtained for $r \in \{\mathrm{I}, \mathrm{II}\}$:

$$Q_{u_t}(\tau|\mathcal{F}_{t-1}) = G_t(\zeta)\, Q^{\mathrm{I}}_{u_t}(\tau|\mathcal{F}_{t-1}) + (1 - G_t(\zeta)) Q^{\mathrm{II}}_{u_t}(\tau|\mathcal{F}_{t-1}) \tag{3.6}$$

$$\begin{aligned} Q^r_{u_t}(\tau|\mathcal{F}_{t-1}) &= \beta^r_0(\tau) + \sum_{i=1}^{p} \beta^r_i Q_{u_{t-i}}(\tau|\mathcal{F}_{t-i-1}) + \sum_{j=1}^{q} \gamma^r_j(\tau)\,|u_{t-j}| \\ &= F_\varepsilon^{-1}(\tau) \theta^{r\,T} z_t^{pq} = \theta^r(\tau)^T z_t^{pq} \end{aligned} \tag{3.7}$$

where the parameters $\beta^r_0(\tau) := \beta^r_0 F_\varepsilon^{-1}(\tau)$ and $\gamma^r_j(\tau) := \gamma^r_j F_\varepsilon^{-1}(\tau)$ are local in a sense that they are dependent on τ and β^r_i are global coefficients operating on the latent past conditional quantiles for all[1] $i \in \mathcal{I}_{1,p}$, $j \in \mathcal{I}_{1,q}$ and $r \in \{I, II\}$. The transition parameters ζ are global as well. This model can be seen as a generalization of the symmetric absolute value CAViaR specification in Engle & Manganelli [2004]. The main difference of CAViaR type models compared to GARCH type models is the direct estimation of the conditional quantiles $Q_{u_t}(\tau|\mathcal{F}_{t-1})$. In what will be discussed in the next session, we estimate the local parameters $\theta(\tau) := \left[\theta^{I}(\tau), \theta^{II}(\tau), \zeta\right] \in \Theta_2^{\tau}$, with Θ_2^{τ} being a compact subset of $\mathbb{R}^{2(p+q+1)+2}$ where $\theta^{r}(\tau) = \left(\beta_0^r(\tau), \beta_1^r(\tau), \dots, \beta_p^r(\tau), \gamma_1^r(\tau), \dots, \gamma_q^r(\tau)\right)^{T}$.

To sum up, asymmetric dynamic behaviour is the result of different parameter regimes rather than a different error distribution in each regime. Since quantiles of the latter are subject to estimation, any symmetric or asymmetric distribution function of ε is permitted for our approach, as long as it satisfies certain regularity properties. A different error disribution would in general be feasible for the threshold model since in this model only one regime at a time is "active". With the smooth transition approach we follow here, identification of an additional error distribution governing a second regime is less clear since the quantile of a linear combination of two innovations is not necessarily the linear combination of the respective quantiles. We will thus leave this for future research.

Estimation Procedure

It is difficult to estimate the CAViaR model, as specified in the previous section, due to the dependence on past conditional quantiles. In this section, we will propose a two-step estimation procedure that is related to the three-stage sieve approximation idea of Xiao & Koenker [2009]. In contrast to their single regime model, we are however faced with the additional complication of estimating parameters from two regimes and the corresponding parametrized (location and scale) transition function. The idea of the estimation procedure is to first approximate the conditional volatility process as defined in equation (3.3) by an ARCH(∞)-approximation and then use the structure of our model defined in equation (3.5) to estimate the CAViaR parameters and the transition parameters in a second stage. While it is assumed that the transition function $G \in \mathfrak{G}$ is known *a priori*, it is not always the case that our objective function is convex in its parameters. In order to estimate the parameters of the model as described above, we therefore employ both quantile regression techniques and an exhaustive machine

[1]For notational convenience the general index set running from $a \in \mathbb{N}$ to $b \in \mathbb{N}$ shall be defined as $\mathcal{I}_{a,b} := (a, \dots, b) \subseteq \mathbb{N}$.

search.

The model defined in (3.6) and (3.7) can be estimated using the objective function

$$\min_{\theta\in\Theta_2^\tau} n^{-1}\sum_{t=1}^{n}\rho_\tau\left(u_t-\theta^{I^T}z_t^{pq}(\theta)G_t(\zeta)+\theta^{II^T}z_t^{pq}(\theta)(1-G_t(\zeta))\right)$$

with $z_t^{pq}(\theta)=(\sigma_{t-1}(\theta),\dots,\sigma_{t-p}(\theta),|u_{t-1}|,\dots,|u_{t-q}|)\in\mathcal{F}_{t-1}$, where $\sigma_t(\theta)$ has the structure defined in (3.3).

The estimation of conditional quantiles would be a linear programming exercise, if it were not for the dependence on the latent conditional volatility process σ_t, which in turn dynamically depends on the parameters θ that have to be estimated. To tackle this issue, we make use of a two-step procedure.

In a first step, each regime's GARCH(p,q) process in equation (3.4) is inverted to ARCH(∞) and estimated using an ARCH(m) representation, where $m\in\mathbb{N}$, in order to find a sieve approximation of $\sigma_t := \sigma_t(z_t,\theta)$. To ensure invertibility, let $A^r(L) := 1-\sum_{i=1}^{p}\beta_i^r L^i$ and $B^r(L) := \sum_{i=0}^{q-1}\gamma_i^r L^i$, where L is the lag-operator such that $u_{t-1}=Lu_t$ for any $t\in\mathcal{I}_{1,n}$.

Assumption 3.2. The polynomials $A^r(L)$ and $B^r(L)$ have no common factors and their roots lie outside the unit disc of the complex plane: for $r\in\{I,II\}$ and $|\phi|\leqslant 1$, it holds that $A^r(\phi)\neq 0$ and $B^r(\phi)\neq 0$.

Hence both GARCH(p,q) regimes defined in equation (3.3) can be inverted separately:

$$A^r(L)\sigma_t^r = B^r(L)\,|u_t| \iff \sigma_t^r = A^{r-1}(L)B^r(L)\,|u_t| = \alpha_0^r+\sum_{j=1}^{\infty}\alpha_j^r\,|u_{t-j-1}|, \tag{3.8}$$

where the coefficients α_j for $j\in\mathcal{I}_{1,m}$ decrease at a geometric rate. Thus there exists constants $b<1$ and c such that $|\alpha_j|<cb^j$. Consequently each conditional volatility regime defined in equation (3.8) can be approximated by a finite dimensional ARCH(m) process such that our model becomes:

$$Q_{u_t}(\tau|\mathcal{F}_{t-1})\approx\left(\alpha_0^I(\tau)+\sum_{j=1}^{m}\alpha_j^I(\tau)\,|u_{t-j-1}|\right)G_t(\zeta)+\left(\alpha_0^{II}(\tau)+\sum_{j=1}^{m}\alpha_j^{II}(\tau)\,|u_{t-j-1}|\right)(1-G_t(\zeta))$$

with $\alpha_j^r(\tau) := \alpha_j^r F_\varepsilon^{-1}(\tau)$ for all $j\in\mathcal{I}_{0,m}$ and $r\in\{I,II\}$ and where the approximation is up to a tight sequence of order b^m as specified above. In order to estimate the conditional volatility process σ_t, we need to identify and estimate $\boldsymbol{\alpha}^r=[\alpha_1^r,\dots,\alpha_m^r]^T$ separately from $F_\varepsilon^{-1}(\tau)$. For this reason, we will not only estimate single conditional quantiles but exploit information from a range of quantiles[2] $(\tau_1, \dots, \tau_K)$ and employ composite quantile regression [Zou & Yuan, 2008] by minimizing the following truncated approximation:

$$\left(\widehat{\boldsymbol{\alpha}}_n^{I}, \widehat{\boldsymbol{\alpha}}_n^{II}, \widehat{\mathbf{q}}_n, \widehat{\zeta}_n\right) = \arg\min_{\alpha \in \boldsymbol{\Theta}_1} \sum_{k=1}^{K} \sum_{t=m+1}^{n} \rho_{\tau_k}\left(u_t - q_k \boldsymbol{\alpha}^{I^T} \boldsymbol{z}_t^m G_t(\zeta) + q_k \boldsymbol{\alpha}^{II^T} \boldsymbol{z}_t^m (1 - G_t(\zeta))\right) \tag{3.9}$$

with $\boldsymbol{z}_t^m = (1, |u_{t-1}|, \dots, |u_{t-m}|)^T$, $\boldsymbol{\alpha}^r = (\alpha_0^r, \dots, \alpha_m^r)^T$ for $r \in \{I, II\}$, $\mathbf{q} = [q_1, \dots, q_K]^T$ where $q_k = F_\varepsilon^{-1}(\tau_k)$ for all $k \in \mathcal{I}_{1,K}$ and $\left[\boldsymbol{\alpha}^{I^T}, \boldsymbol{\alpha}^{II^T}, \mathbf{q}^T, \zeta^T\right]^T \in \boldsymbol{\Theta}_1$, which is assumed to be a compact subset of $\mathbb{R}^{2(m+1)+K+2}$. The function $\rho_\tau(u) = u\left(\tau - \mathbb{1}_{\{u<0\}}\right)$ denotes the quantile loss function.

In the second step, we can now use equation (3.8) to obtain an expression for

$$\sigma_t(\widehat{\boldsymbol{\alpha}}_n) = \left(\widehat{\boldsymbol{\alpha}_n^{I}}^T \boldsymbol{z}_t^m\right) G_t\left(\widehat{\zeta}_n\right) + \left(\widehat{\boldsymbol{\alpha}_n^{II}}^T \boldsymbol{z}_t^m\right)\left(1 - G_t\left(\widehat{\zeta}_n\right)\right). \tag{3.10}$$

and we can estimate the CAViaR model according to equation (3.6) and (3.7) for a single quantile by minimizing

$$\widehat{\theta}_n(\tau) = \arg\min_{\theta \in \boldsymbol{\Theta}_2^\tau} \sum_{t=t_0}^{n} \rho_\tau\left(u_t - \theta^{I^T} \boldsymbol{z}_t^{pq}(\widehat{\boldsymbol{\alpha}}_n) G_t(\zeta) + \theta^{II^T} \boldsymbol{z}_t^{pq}(\widehat{\boldsymbol{\alpha}}_n)\,(1 - G_t(\zeta))\right), \tag{3.11}$$

with $\theta(\tau) = \left[\theta^{I^T}(\tau), \theta^{II^T}(\tau), \zeta(\tau)\right]^T \in \boldsymbol{\Theta}_2^\tau$, $t_0 = [(m+p) \vee q] + 1$ and z^{pq} as defined above. For given location and scale parameters, this transformation allows estimation using standard quantile autoregression techniques [Koenker & Zhao, 1996]. Note that, we do not make use of the estimated quantiles $\widehat{\mathbf{q}}_n$ from the first stage, and also re-estimate, the location and scale parameters, which we denote by $\zeta(\tau)$ in the second stage to indicate the estimation jointly with the local parameters.

Neither of the objective functions (3.9) and (3.11) are convex in the scale parameter. In addition to this, the quantile loss function is not differentiable which further complicates our analysis. For this reason we have to use a grid search over the space of feasible scale parameters in the first stage in which we use a smoothed version of the objective function ρ which we define as $\rho^*(u) = \rho(u)$ if $|u| > \delta$ and u^2/δ otherwise, with smoothing parameter δ. This approach is commonly used in the literature (see e.g. Huber [1964]; Zheng [2011]) and allows us to estimate the parameters for a given gridpoint using gradient based methods. For the second stage we use the standard quantile loss function ρ combined with a grid search over all feasible location/scale pairs. We define feasible pairs to satisfy the assumption that we do in fact observe

[2] Xiao & Koenker [2009] solve this by first estimating the parameters for each τ and then exploit their structure $\alpha_j^r(\tau) := \alpha_j^r F_\varepsilon^{-1}(\tau)$ in additional minimum distance estimation step.

two regimes in our sample and will informally denote this subspace as $\mathfrak{Z}$, which will be made explicit in the empirical section. The reason that we use a two dimensional grid in the second stage, as opposed to the one dimensional grid in the first stage is that we exploit the fact that for a given location/scale pair we have a linear quantile regression problem which we can estimate directly by employing a standard interior point method. It is instructive to summarize the algorithm described above using the following pseudo-code:

$(l^1_{\min}, \widehat{\alpha}^{I}_n, \widehat{\alpha}^{II}_n, \widehat{\zeta}_n, \widehat{\eta}_n) \leftarrow (\infty, 0, 0, 0, 0)$
for all $\eta \in \{\eta_1, \dots, \eta_{k_\eta}\}$ **do**
Define $G(\xi_t, \zeta, \eta)$ using $\xi_t \leftarrow \xi(z_t)$ for given scale η as a function of ζ
Estimate $\widehat{\alpha}^{I}_{n,k_\eta}, \widehat{\alpha}^{II}_{n,k_\eta}, \widehat{q}_{k_\eta}, \widehat{\zeta}_{k_\eta}$ and obtain loss $l^1_{k_\eta}$ according to (3.9) by smoothed CRQ
if $l^1_{k_\eta} \leqslant l^1_{\min}$ **then**
$(l^1_{\min}, \widehat{\alpha}^{I}_n, \widehat{\alpha}^{I}_n, \widehat{\zeta}_n, \widehat{\eta}_n) \leftarrow (l^1_{k_\eta}, \widehat{\alpha}^{I}_{n,k_\eta}, \widehat{\alpha}^{II}_{n,k_\eta}, \widehat{\zeta}_{k_\eta}, \eta)$
end if
end for
Calculate $\widehat{\sigma}_t$ according to equation (3.10) using $\widehat{\boldsymbol{\alpha}}_n$
Construct $z^{pq}_t = (\widehat{\sigma}_{t-1}, \dots, \widehat{\sigma}_{t-p-1}, |u_{t-1}|, \dots, |u_{t-q-1}|)$
$(l^2_{\min}, \widehat{\theta}^{I}_n, \widehat{\theta}^{II}_n, \widehat{\zeta}_n, \widehat{\eta}_n) \leftarrow (\infty, 0, 0, 0, 0)$
for all $(\zeta, \eta) \in \{\zeta_1, \dots \zeta_{k_\zeta}\} \times \{\eta_1, \dots, \eta_{k_\eta}\} \cap \mathfrak{Z}$ **do**
Calculate $G(\xi_t, \zeta, \eta)$ using $\xi_t \leftarrow \xi(z_t)$ for given location ζ and scale η
Estimate $\widehat{\theta}^{I}_{n,k_{\zeta,\eta}}, \widehat{\theta}^{II}_{n,k_{\zeta,\eta}}$ and obtain loss $l^2_{k_{\zeta,\eta}}$ according to (3.11) by linear RQ
if $l^2_{k_{\zeta,\eta}} \leqslant l^2_{\min}$ **then**
$(l^2_{\min}, \widehat{\theta}^{I}_n, \widehat{\theta}^{I}_n, \widehat{\zeta}_n, \widehat{\eta}_n) \leftarrow (l^2_{k_{\zeta,\eta}}, \widehat{\theta}^{I}_{n,k_{\zeta,\eta}}, \widehat{\theta}^{II}_{n,k_{\zeta,\eta}}, \zeta, \eta)$
end if
end for

Algorithm 3.1: Two-stage estimation procedure

What remains to be addressed is the selection of K and the specification of $(\tau_1, \dots, \tau_K)$. Here we face a trade-off. On the one hand, we would like to learn from as many quantiles of the distribution of $u_t|\mathcal{F}_{t-1}$ as possible. On the other hand, the global parameters θ_j for $j \in \mathcal{I}_{1,2(p+q+1)}$ are not identified at the median due to the model structure $\theta_j(\tau) = F^{-1}_{\varepsilon}(\tau)\theta_j$. For finite samples we will thus introduce extra noise if we include quantiles around $\tau = 0.5$, due to this "weak" identification problem. While a data-driven optimal selection of a vector of τ's would be feasible, this goes beyond the scope of this paper. We refer the interested reader to Zhao & Xiao [2014]. In the Monte Carlo section we will however present the results of experiments with different choices of δ, where δ is the width of the window in the vector of τ's around the median that is left out for estimation.

Asymptotic Results

In this section, the two-stage estimation procedure is shown to yield consistent and asymptotically normal estimates for the proposed ANST-GACQ model. Throughout this section it is assumed that in addition to the previously defined assumptions the following statements hold:

Assumption 3.3. The errors ε_t are i.i.d. distributed with zero mean and finite variance $\sigma^2 = \mathrm{Var}(\varepsilon_t) < +\infty$. Its distribution function $F_\varepsilon(x) \in C^1(\mathbb{R})$ has strictly positive density $f_\varepsilon(x)$ at $F_\varepsilon^{-1}(\tau_k)$ for all $k \in \mathcal{I}_{1,K}$, which is uniformly bounded by a finite constant M and Lipschitz continuous.

Assumption 3.4. The conditional distribution function $F_{u_t|\mathcal{F}_{t-1}}(x) \in C^1(\mathbb{R})$ has strictly positive density $f_{u_t|\mathcal{F}_{t-1}}(x)$ at $F^{-1}_{u_t|\mathcal{F}_{t-1}}(\tau_k)$ for all $k \in \mathcal{I}_{1,K}$, which is uniformly bounded with finite constant M and Lipschitz continuous.

Assumption 3.5. There exists a small positive constant $\delta > 0$ such that $\mathbb{E}\left|u_t G_t(\zeta_0)\right|^{2+\delta} < +\infty$, $\mathbb{E}\left|u_t\right|^{2+\delta} < +\infty$ and $\mathbb{E}\left|u_t \frac{\partial G_t(\zeta_0)}{\partial \zeta}\right|^{2+\delta} < +\infty$. In addition to this, u_t is β-mixing with mixing coefficient satisfying $\beta_s \to 0$ as $s \to \infty$ with $m \leqslant s^{1/4}$, $\sum_{s=1}^{\infty} \beta_s^{\delta/(2+\delta)} < +\infty$, $\sum_{s=1}^{\infty} \beta_s s^{2/\delta} < +\infty$ where $\beta_s = \beta_0 s^{-(2+\delta)}$.

Assumption 3.6. Let $\boldsymbol{a}^\mathsf{T} = [\boldsymbol{\alpha}^{\mathrm{I},\mathsf{T}}, \boldsymbol{\alpha}^{\mathrm{II},\mathsf{T}}, \mathbf{q}^\mathsf{T}, \boldsymbol{\zeta}^\mathsf{T}]^\mathsf{T}$ in $\boldsymbol{x}_{t,k}(\boldsymbol{a})$ which is defined in equation (3.14). The matrix

$$\mathbf{D}_{1,m,n}(\boldsymbol{a}) := \mathbb{E}\left[n^{-1} \sum_{t=m}^{n} \left[\sum_{k=1}^{K} \boldsymbol{x}_{t,k}(\boldsymbol{a}) \boldsymbol{x}_{t,k}(\boldsymbol{a})^\mathsf{T}\right] / \sigma_t\right]$$

evaluated at $\boldsymbol{a}_0$ has minimum and maximum eigenvalues denoted by $\lambda_{n,\min}$ and $\lambda_{n,\max}$ satisfying $\lim_{n\to\infty} \inf \lambda_{n,\min} > 0$ and $\lim_{n\to\infty} \sup \lambda_{n,\max} < +\infty$. In addition to this we assume that $\mathbb{E}[G_t(\zeta)\, \boldsymbol{z}_t^m, (1 - G_t(\zeta))\boldsymbol{z}_t^m, \boldsymbol{z}_t^m(G_t(\zeta) - G_t(\zeta_0))]^\mathsf{T}[G_t(\zeta)\, \boldsymbol{z}_t^m, (1 - G_t(\zeta))\boldsymbol{z}_t^m, \boldsymbol{z}_t^m(G_t(\zeta) - G_t(\zeta_0))]$ has full rank for any $\zeta \neq \zeta_0$.

Assumption 3.7. The number of lags for the ARCH(m)-approximation satisfies both $\log(n)/m \to 0$ and $mn^{-\frac{1}{2}} \to 0$.

Assumption 3.8. There exist small positive constants $\delta_1 > 0$ and $\delta_2 > 0$ such that

$$\mathbf{P}\left(\max_{1\leqslant t\leqslant n} u_t^2 > n^{\delta_1}\right) \leqslant \exp\left(-n^{\delta_2}\right).$$

Assumption 3.2-3.5 and the fact that the transition function has range $\mathcal{R} \subseteq [0,1]$ ensure that the process u_t is stationary and weakly dependent. The moment assumptions on u_t, $u_t G_t(\zeta_0)$ and $u_t \partial G_t(\zeta)/\partial\zeta$ are required for a weakly dependent central limit theorems. We assume with out loss of generality that the median of ε is normalized to zero such that $F_\varepsilon^{-1}(1/2) = 0$. A detailed discussion of mixing-requirements of regime switching models can be found in Carrasco & Chen [2002]; Meitz & Saikkonen [2008] and references therein. Note that a sufficient condition for the existence of the finite $(2+\delta)$ moment of the term $u_t \partial G_t(\zeta)/\partial\zeta$ is the existence of $(4+\delta)$ moments for u_t and $\mathbb{E}\left|\partial G_t(\zeta)/\partial\zeta\right| < \infty$. If we have exogenous switching, i.e. if ξ_t is a different time series, the existence of $(2+\delta)$ finite moments of u_t suffices. Assumption 3.6 is an identification assumption which ensures that the two regimes have different conditional volatility processes, that the data in the two both regimes are not perfectly correlated and that for the linear transition function, which has slope zero on subsets of its domain, that with positive probability there is data in both regimes. Assumption 3.7 restricts the rate of the ARCH(m) approximation, ensuring that we have a sufficient number of lags, which controls the approximation error. In our empirical application we will chose $m = cn^{1/4}$ for some positive constant $c > 0$. Finally, Assumption 3.8 is a technical regularity condition that is needed for the sieve estimation in the first stage.

We will now show the asymptotic properties of our proposed estimation procedure. First, we will show that the sieve approximation of both regimes' underlying GARCH processes holds and the approximation error due to the m^{th}-order truncation is bounded in probability.

Theorem 3.1 (Identification and First Stage Consistency). Let $\boldsymbol{a}^{\mathsf{T}} = [\boldsymbol{\alpha}^{\mathrm{I},\mathsf{T}}, \boldsymbol{\alpha}^{\mathrm{II},\mathsf{T}}, \mathbf{q}^{\mathsf{T}}, \boldsymbol{\zeta}^{\mathsf{T}}]^{\mathsf{T}}$. Under Assumptions 3.1-3.8, it holds for $n \to \infty$

$$\|\widehat{\boldsymbol{a}}_n - \boldsymbol{a}_0\| = \mathcal{O}_p\left(\frac{m}{n}\right).$$

The following asymptotic normality result of the first stage estimator provides a preliminary result for the asymptotic characteristics of the interim volatility estimator and the second stage quantile estimator.

Corollary 3.1 (First Stage Bahadur Representation). Let $s_k = f_\varepsilon(F_\varepsilon^{-1}(\tau_k))$, $\mathbf{s}^{\mathsf{T}} = [s_1, \ldots, s_K]$, $\mathbf{q}^{\mathsf{T}} = [q_1, \ldots, q_K]$. Also let $\boldsymbol{\alpha}^{\Delta} = \boldsymbol{\alpha}^{\mathrm{I}} - \boldsymbol{\alpha}^{\mathrm{II}}$ and $\bar{\boldsymbol{\alpha}}(\zeta) = G_t(\zeta)\,\boldsymbol{\alpha}^{\mathrm{I}} + (1 - G_t(\zeta))\boldsymbol{\alpha}^{\mathrm{II}}$. Then under

Assumptions 3.1-3.8, it holds for $n \to \infty$

$$\sqrt{n}\begin{bmatrix} \widehat{\alpha}_n^{I} - \alpha_0^{I} \\ \widehat{\alpha}_n^{II} - \alpha_0^{II} \\ \widehat{\zeta}_n - \zeta_0 \end{bmatrix} \approx \left(\sum_{k=1}^{K} s_k q_k^2\right)^{-1} \mathbf{D}_{n,m}^{-1} \frac{1}{\sqrt{n}} \sum_{t=m+1}^{N} \begin{bmatrix} G_t(\zeta)\, z_t^m \\ 1 - G_t(\zeta)\, z_t^m \\ z_t^{m,T} \alpha_0^{\Delta} \frac{\partial G_t(\zeta)}{\partial \zeta} \end{bmatrix} \sum_{k=1}^{K} q_k \left(\mathbb{1}\left\{u_t \leqslant F^{-1}_{u_t|\mathcal{F}_{t-1}}(\tau_k)\right\} - \tau_k\right)$$

where the approximation is up to a stochastically negligible sequence of order $(m/n)^{1/2}$ and where

$$\mathbf{D}_{n,m} := -\mathbb{E}\frac{1}{n}\sum_{t=m}^{n}\frac{1}{\sigma_t}\begin{bmatrix} \mathbf{\Omega}_{11} & \mathbf{\Omega}_{12} \\ \mathbf{\Omega}_{12}^T & \mathbf{\Omega}_{22} \end{bmatrix}.$$

with respective blocks

$$\mathbf{\Omega}_{11} = \begin{bmatrix} G_t(\zeta_0)^2 & G_t(\zeta_0)\,(1 - G_t(\zeta_0)) \\ G_t(\zeta_0)\,(1 - G_t(\zeta_0)) & (1 - G_t(\zeta_0))^2 \end{bmatrix} \otimes z_t^m z_t^{m,T},$$

$$\mathbf{\Omega}_{12} = z_t^{mT} \alpha_0^{\Delta} \left(\begin{bmatrix} G_t(\zeta_0) \\ 1 - G_t(\zeta_0) \end{bmatrix} \otimes z_t^m \frac{\partial G}{\partial \zeta_0^T}\right),$$

$$\mathbf{\Omega}_{22} = (z_t^{mT} \alpha_0^{\Delta})^2 \frac{\partial G}{\partial \zeta_0}\frac{\partial G}{\partial \zeta_0^T}.$$

This completes the first step sieve estimation, concerning the estimates for α, determining the latent process σ_t defined according to equation (3.10) which we need for the second step of the estimation. In addition to this we get an asymptotic distribution for this first stage, which will enter the final asymptotic expression. Applying quantile regression techniques with

$$z_t^{pq} = (1, \widehat{\sigma}_{t-1}, \ldots, \widehat{\sigma}_{t-p-1}, |u_{t-1}|, \ldots, |u_{t-q-1}|)$$

we can then estimate the CAViaR parameters $\theta(\tau)$. The asymptotic theory for the second stage involves a non-differentiable objective function and the dependence of the second stage on the first stage parameters (combination of a Type II an Type IV empirical process problems in Andrews [1994]). The following two theorems provide consistency and asymptotic normality results of the final ANST-GACQ estimator using the preliminary results from the first stage.

Theorem 3.2 (Second Stage Consistency)**.** Under Assumptions 3.1-3.8, the second-stage estimator is $\sqrt{n}$-consistent, that is, for $n \to \infty$ and $\tau \in (0,1)$

$$\left\|\widehat{\theta}_n(\tau) - \theta_0(\tau)\right\| = \mathcal{O}_p(n^{-\frac{1}{2}}).$$

Theorem 3.3 (Second Stage Asymptotic Normality)**.** If Assumptions 3.1-3.8 hold, the second-stage estimates $\widehat{\theta}_n(\tau)$ are asymptotically normal

$$\sqrt{n}\left(\widehat{\theta}_n(\tau)-\theta_0(\tau)\right) \rightsquigarrow \mathcal{N}\left(0, \boldsymbol{\Gamma}_{\theta,0}^{-1}\mathbb{E}\left[\mathbf{M}_t\Xi^{\tau}\mathbf{M}_t\right]\boldsymbol{\Gamma}_{\theta,0}^{-1}\right),$$

where

$$\mathbf{M}_t = \mathbf{D}_2\left[z_t(\alpha_0)G_t(\zeta_0), z_t(\alpha_0)(1-G_t(\zeta_0)), \theta_0^{\Delta T}z_t(\alpha_0)\frac{\partial G_t(\zeta_0)}{\partial\zeta}\right. \tag{3.12}$$
$$\left., z_t^m G_t(\zeta_0), z_t^m(1-G_t(\zeta_0)), \alpha_0^{\Delta} z_t^m \frac{\partial G_t(\zeta_0)}{\partial \zeta}\right],$$

$\mathbf{D}_2 = \left[\mathbf{I}_{2(p+q+1)}, \frac{1}{\sum_{k=1}^{K} s_k q_k^2}\Gamma_{\alpha,0}\mathbf{D}_n^{-1}\right]$, and with a typical element of Ξ^{τ} defined as

$$\Xi_{i,j} = \begin{cases} \tau(1-\tau)/f_{\varepsilon}(F_{\varepsilon}^{-1}(\tau))^2 & \text{if } i \leqslant p+q+1 \text{ and } j \leqslant p+q+1 \\ q_i q_j(\tau_i \wedge \tau_j)(1-\tau_i \vee \tau_j)/[f_{\varepsilon}(F_{\varepsilon}^{-1}(\tau_i))f_{\varepsilon}(F_{\varepsilon}^{-1}(\tau_j))] & \text{if } i > p+q+1 \text{ and } j > p+q+1 \\ q_i(\tau_i \wedge \tau)(1-\tau_i \vee \tau)/[f_{\varepsilon}(F_{\varepsilon}^{-1}(\tau_i))f_{\varepsilon}(F_{\varepsilon}^{-1}(\tau))] & \text{otherwise } (i > j \text{ w.l.o.g}). \end{cases} \tag{3.13}$$

Having shown the necessary theoretical properties of the estimation procedure, the model is now confronted with data in order to see how it works empirically compared to its single-regime VaR and regime-switching conditional variance counterparts.

Monte-Carlo Simulations

The following Monte Carlo study is divided into two main parts. First, our proposed asymmetric non-linear smooth transition generalized conditional quantile (ANST-GACQ) procedure will be analyzed with respect to different specifications and estimation parameters, i.e. different sample sizes, the amount of lags used to approximate the first stage ARCH(∞) model or the width δ around the median that is excluded from the composite quantile estimation due to non-identification of the central tendency of the process. Results are then contrasted to the regime switching GARCH model of Anderson et al. [1999] for a range of error distributions in a second step. After this, the consequences of misspecifying the transition function are shown, and finally, the robustness with respect to outliers of these two models is investigated. All experiments are conducted using Ox [Doornik, 2009] with extensions written in C for the computationally more intensive parts. The length of the considered time series is set to $n = 1{,}000$, the number of simulations per experiment is $s = 100$, we estimate a range of $k = 9$ quantiles with $\delta = 0.25$ and $(\tau_1, \tau_K) = (0.05, 0.95)$, the truncation parameter for the ARCH approximation is set to $m = \lceil \frac{3}{2} n^{\frac{1}{4}} \rceil$ and the grid size to

$(k_\zeta, k_\eta) = (30, 30)$. This is kept constant for all subsequent experiments, unless explicitly defined otherwise. The true global parameter vector for both processes is chosen to be $\theta_0 = (\beta_0^{\mathrm{I}}, \beta_1^{\mathrm{I}}, \gamma_1^{\mathrm{I}}, \beta_0^{\mathrm{II}}, \beta_1^{\mathrm{II}}, \gamma_1^{\mathrm{II}})_0 = (0.50, 0.15, 0.60, 0.25, 0.30, 0.15)$ and the location/scale parameter pair equals $\zeta_0 = (\zeta, \eta)_0 = (0.00, 0.2)$. This implies that the regimes' unconditional variances defined by $\beta_0^r/(1 - \beta_1^r - \gamma_1^r)$ for $r \in \{\mathrm{I}, \mathrm{II}\}$ are given by 2 and 0.45, respectively. In addition to this, the parameters are chosen such that they satisfy the invertibility conditions of both regimes' GARCH processes. Figure 3.1 shows all three considered transition functions, evaluated at their true parameters, on the domain of a typical realization of process $\{u_t\}$.

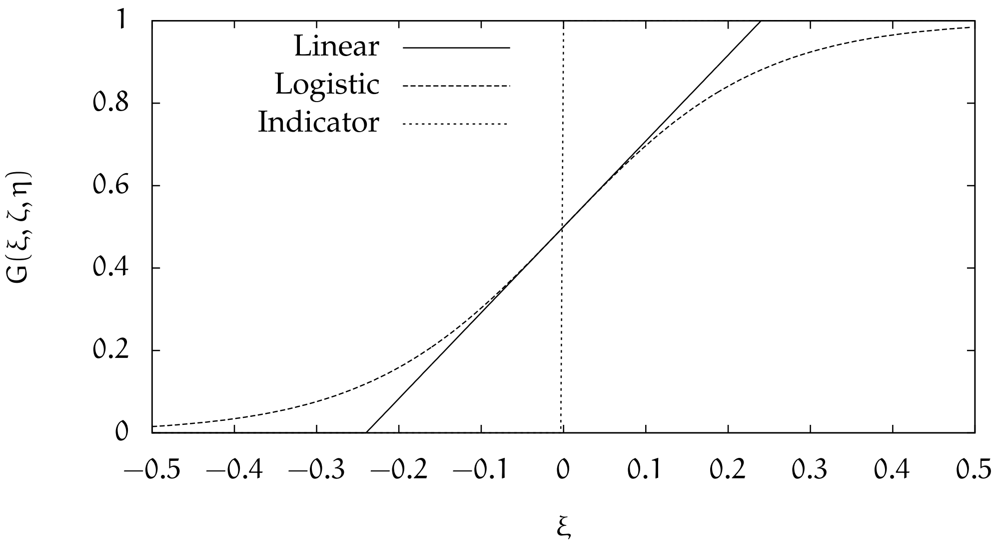

Figure 3.1: Logistic, Linear and Threshold Transition Function on $[F_{\xi,t}^{-1}(0.05), F_{\xi,t}^{-1}(0.95)]$

The second part of the experiment focuses on the comparison of the new procedure with respect to previously established models. The main measure of performance will rather be the (absolute) prediction error as averaged over the sample and the coverage ratio of the estimated 5% value at risk, reported as *M(A)PE* and *coverage*, respectively. The coverage ratio is the proportion of observations that falls below the estimated value at risk. By definition, this ratio should be close to τ, for the τ% value at risk.

In what follows we restrict the grid for both location and scale in the following way. The location must satisfy $\zeta \in [\underline{\zeta}, \bar{\zeta}]$ with unconditional sample quantiles $\underline{\zeta} = \widehat{F}_{u_t}^{-1}(0.1)$ and $\bar{\zeta} = \widehat{F}_{u_t}^{-1}(0.9)$ for all transition functions. The scale is restricted to $\eta \in [\underline{\eta}, \bar{\eta}(\underline{\zeta}, \bar{\zeta})]$ with fixed $\underline{\eta} = 10^{-3}$ and $\bar{\eta}(\underline{\zeta}, \bar{\zeta}) = \bar{\zeta} - \underline{\zeta}$ for the linear transition function and $\bar{\eta}(\underline{\zeta}, \bar{\zeta}) = \left[\log(0.1^{-1} - 1)(0.5\bar{\zeta} - 0.5\underline{\zeta})\right]^{-1}$ for the logistic transition function. The former restriction for the linear transition function ensures that the cut-off points for its lower and upper bounds coincide with the minimum and maximum location, respectively, when the location is centered. The latter represents the inverse of the logistic function with respect to the scale evaluated at 0.1 and location half-way between the 10^{th} and 90^{th} unconditional quantile i.e. the center of the considered location grid. As opposed to the linear one, the logistic density is non-zero on the whole domain $\mathbb{R}$. Thus the condition has to be relaxed such that only those scales are considered which are small enough to ensure that at most 10% of the observations are below $\underline{\zeta}$. In addition to this, the following triangular joint restriction must hold: $2\bar{\eta}(\underline{\zeta},\bar{\zeta})(\bar{\zeta}-\underline{\zeta})^{-1}(\bar{\zeta}-\zeta) \leqslant 0$ and $2\bar{\eta}(\underline{\zeta},\bar{\zeta})(\bar{\zeta}-\underline{\zeta})^{-1}(\zeta-\underline{\zeta}) \geqslant 0$. This ensures that, as the location gets close to its minimum or maximum, respectively, the scales must become smaller as well, in order to avoid location/scale pairs that would suggest that there is only one regime. The whole set of restrictions can be summarized by defining the constraint set

$$\begin{aligned}\mathfrak{Z} := \big\{(\eta,\zeta) \in \mathbb{R}\times\mathbb{R}_+ : \zeta \in [\underline{\zeta},\bar{\zeta}],\ \eta \in [\underline{\eta},\bar{\eta}(\underline{\zeta},\bar{\zeta})],\\ 2\bar{\eta}(\underline{\zeta},\bar{\zeta})(\bar{\zeta}-\underline{\zeta})^{-1}(\bar{\zeta}-\zeta) \leqslant 0,\\ 2\bar{\eta}(\underline{\zeta},\bar{\zeta})(\bar{\zeta}-\underline{\zeta})^{-1}(\zeta-\underline{\zeta}) \geqslant 0\big\}.\end{aligned}$$

As a first experiment we consider how our estimates improve with the sample size. We estimate the model with $N = 1000$, $N = 2000$ and $N = 4000$ observations, respectively. Table 3.1 summarizes the result.

It is reassuring that the RSME's of the parameter estimates decrease as the sample gets larger. Note that even while the estimates for the paramters β_1 have larger RMSE's, which is not surprising due to the fact that the conditional volatility process is predicted in the first stage, they are decreasing with respect to sample size as well. Although the second-stage transition parameters are estimated more precisely as the sample size increases, RMSE's seem to go down slower than expected. This might be due to the fact that they enter the model non-linearly through the transition function, and are thus difficult to estimate from a numerical point of view, particularly in the first stage where we use a smooth approximation of the quantile loss function, which is often flat around the true parameters[3]. Finally, mean absolute prediction errors are decreasing and, unsurprisingly, coverage ratios are perfectly accurate by definition due

[3]We experience occasional issues with no line-search improvement depending on the amount of smoothing.

	N = 1000		N = 2000		N = 4000	
	Bias	RMSE	Bias	RMSE	Bias	RMSE
β_0^I	0.1797	0.4120	0.1223	0.2489	0.0749	0.2123
β_1^I	-0.0129	0.3559	0.0065	0.2779	-0.0068	0.1179
γ_1^I	-0.1196	0.3270	-0.0653	0.2276	-0.0228	0.1404
β_0^{II}	0.0319	0.1650	0.0337	0.1231	0.0236	0.1006
β_1^{II}	0.0263	0.2274	0.0272	0.1701	0.0187	0.1021
γ_1^{II}	0.0050	0.1310	0.0011	0.0984	0.0150	0.0891
ζ	0.3798	0.4923	0.3252	0.4703	0.4595	0.5632
$\zeta(\tau)$	0.1063	0.3730	0.1149	0.3138	0.0541	0.2547
η	0.0870	0.1341	0.0627	0.1170	0.0743	0.1229
$\eta(\tau)$	-0.0622	0.0867	-0.0327	0.0807	-0.0353	0.0826
MPE	0.0082	0.0482	0.0027	0.0447	0.0032	0.0203
MAPE	0.1259	0.1310	0.0974	0.1004	0.0682	0.0700
coverage	0.0008	0.0013	0.0004	0.0007	0.0001	0.0003

Table 3.1: Sample Size

to the way the quantile regression estimator is constructed[4].

In order to study the influence of the amount of lags in the ARCH(m) approximation, in a next experiment we consider different multiples of $n^{\frac{1}{4}}$, which is the order of the rate that needs to be satisfied according to the theory. The results are reported in in Table 3.2. We can see that the results are fairly constant with respect to our approximation parameter, although there is a slight U-shape with the optimum in terms of mean absolute prediction errors between $c = 1.0$ and $c = 2.0$ which corresponds to 6 and 12 lags in our $N = 1000$ sample, respectively. We have thus no reason to deviate from the proposed $c = 1.5$ in Xiao & Koenker [2009].

Table 3.3 displays results for different values of δ, which is the width of the window left out around the median, in our first stage estimation. While according to theory δ should tend towards zero as the sample size increases, for finite samples we face the trade-off of introducing noise to our estimates that stems from the non-identification of the global parameters around the median, and on the other hand not considering most of the observations that naturally occur around the median, which also increases our variance.

As we can see, the precision of the estimates seems to be rather stable with respect

[4]The reason we report the coverage ratio, is to allow for a direct comparison to the ANST-GARCH models in the second part of the experiment, for which this property does not necessarily hold.

	$c = 1.0$		$c = 2.0$		$c = 3.0$		$c = 3.5$	
	Bias	RMSE	Bias	RMSE	Bias	RMSE	Bias	RMSE
β_0^I	0.1444	0.4107	0.1472	0.4023	0.1487	0.4076	0.2107	0.4580
β_1^I	0.0402	0.4108	0.0254	0.3986	0.0454	0.4821	0.0643	0.4523
γ_1^I	-0.1186	0.3157	-0.1184	0.3085	-0.1138	0.3685	-0.1816	0.3700
β_0^{II}	0.0033	0.1714	0.0187	0.1695	0.0308	0.1688	0.0353	0.1644
β_1^{II}	0.0250	0.2224	0.0227	0.2242	0.0232	0.2383	0.0205	0.2216
γ_1^{II}	0.0167	0.1466	0.0011	0.1384	-0.0032	0.1383	-0.0008	0.1340
ζ	0.2025	0.5823	0.2447	0.6022	0.3327	0.5587	0.4220	0.6611
$\zeta(\tau)$	0.1062	0.3853	0.0955	0.3861	0.1439	0.4052	0.1829	0.3983
η	0.0682	0.1502	0.0419	0.1381	0.0084	0.1203	0.0458	0.1427
$\eta(\tau)$	-0.0565	0.0851	-0.0659	0.0834	-0.0559	0.0839	-0.0561	0.0871
MPE	0.0103	0.0512	0.0118	0.0523	0.0054	0.0638	0.0068	0.0515
MAPE	0.1245	0.1298	0.1245	0.1294	0.1356	0.1435	0.1317	0.1373
coverage	0.0006	0.0014	0.0006	0.0012	0.0008	0.0014	0.0009	0.0014

Table 3.2: Order of Approximation: $m = \lceil cn^{1/4} \rceil$

	$\delta = 0.15$		$\delta = 0.20$		$\delta = 0.30$		$\delta = 0.50$	
	Bias	RMSE	Bias	RMSE	Bias	RMSE	Bias	RMSE
β_0^I	0.1689	0.4320	0.1830	0.4597	0.1099	0.3801	0.1641	0.4591
β_1^I	0.0367	0.4247	-0.0181	0.4362	0.0481	0.4456	0.0298	0.4371
γ_1^I	-0.1364	0.3431	-0.1167	0.3195	-0.1036	0.2806	-0.1239	0.3428
β_0^{II}	0.0129	0.1879	0.0247	0.1586	0.0259	0.1891	0.0266	0.1698
β_1^{II}	0.0412	0.2289	0.0280	0.2134	0.0302	0.2492	0.0563	0.2574
γ_1^{II}	-0.0006	0.1298	-0.0022	0.1199	-0.0048	0.1406	-0.0183	0.1148
ζ	0.1913	0.5503	0.2884	0.5895	0.2226	0.5715	0.2324	0.5871
$\zeta(\tau)$	0.1362	0.4059	0.1326	0.3895	0.1142	0.3804	0.1597	0.3872
η	0.0537	0.1453	0.0476	0.1396	0.0519	0.1370	0.0623	0.1463
$\eta(\tau)$	-0.0611	0.0879	-0.0630	0.0857	-0.0624	0.0855	-0.0597	0.0845
MPE	0.0104	0.0514	0.0152	0.0521	0.0115	0.0533	0.0092	0.0521
MAPE	0.1276	0.1329	0.1242	0.1301	0.1244	0.1292	0.1268	0.1329
coverage	0.0008	0.0014	0.0008	0.0013	0.0006	0.0012	0.0008	0.0012

Table 3.3: $\delta \in (0,1)$ where central $0.5 \pm \frac{\delta}{2}$ quantiles are not estimated

to different values of δ as well. For the remainder of the simulations and the empirical application we choose $\delta = 0.25$, which concludes the first part of the Monte Carlo study.

In the following sequence of experiments we compare estimates of conditional quantiles with the traditional approach of using GARCH estimates for conditional variances from which the value at risk is calculated using the quantile of the specified innovation distribution. We consider smooth transition GARCH models specified with both Normal and Student's t error as a reference and contrast it to our smooth transition GACQ estimator. Note that the first two models are estimated using maximum likelihood, with location and scale parameters estimated jointly with the other parameters such that local and global location and scale estimates, respectively, are equal by definition whereas the GACQ estimator re-estimates them in a second stage. Variances are normalized to one for all distributions, such that we can immediately compare the estimates.

We start with a data generating process that exhibits standard normal innovations. GARCH-N refers to the correctly specified GARCH model with normal errors and GARCH-t to the misspecified GARCH model assuming Student errors. As opposed to these models, our proposed GACQ estimator does not require a distributional assumption.

	GACQ		GARCH-N		GARCH-t	
	Bias	RMSE	Bias	RMSE	Bias	RMSE
β_0^I	0.1444	0.4107	0.0343	0.0986	-0.1489	0.2789
β_1^I	0.0402	0.4108	-0.1351	0.1410	-0.1020	0.1405
γ_1^I	-0.1186	0.3157	-0.1507	0.1752	-0.0145	0.2500
β_0^{II}	0.0033	0.1714	-0.0820	0.1007	0.0180	0.2037
β_1^{II}	0.0250	0.2224	-0.0190	0.0991	-0.0938	0.1928
γ_1^{II}	0.0167	0.1466	-0.0571	0.1094	0.1146	0.2908
ζ	0.2025	0.5823	0.1248	0.3144	0.4136	0.6883
$\zeta(\tau)$	0.1062	0.3853	0.1248	0.3144	0.4136	0.6883
η	0.0682	0.1502	-0.1478	0.1593	73247	248720
$\eta(\tau)$	-0.0565	0.0851	-0.1478	0.1593	73247	248720
MPE	0.0103	0.0512	0.0771	0.0861	-0.2142	0.3402
MAPE	0.1245	0.1298	0.1461	0.1568	0.3320	0.3877
coverage	0.0006	0.0014	0.0071	0.0094	-0.0150	0.0219

Table 3.4: Model comparison: GACQ and GARCH with Normal errors

It comes with no surprise that the correctly specified GARCH maximum likelihood estimator yields the best parameter estimates for this DGP. In terms of parameter estimates, the GARCH-t model is also performing rather well. However, when it comes to calculating the conditional quantiles, the wrong assumption about the innovation distribution has serious negative consequences on the prediction errors, as can be seen looking at the GARCH-t estimates in Table 3.4. It can also be seen that the maximum likelihood estimates for the scale are severely biased which likely stems from the fact that the estimation procedure is numerically rather unstable. Of course our distribution agnostic GACQ model comes with the price of an efficiency loss in the parameter estimates, however due to more precise location and scale estimate it performs rather well compared to the correctly specified GARCH-N model in terms of prediction errors.

	GACQ		GARCH-N		GARCH-t	
	Bias	RMSE	Bias	RMSE	Bias	RMSE
β_0^I	0.1936	0.6391	0.0209	0.1717	0.0298	0.1314
β_1^I	0.0817	0.8266	-0.0723	0.1494	-0.1037	0.1293
γ_1^I	-0.1550	0.4665	-0.2043	0.2504	-0.1763	0.2143
β_0^{II}	0.0391	0.2259	-0.0824	0.1638	-0.1066	0.1271
β_1^{II}	0.0288	0.3015	-0.0279	0.1677	-0.0270	0.1181
γ_1^{II}	-0.0007	0.1604	-0.0749	0.1085	-0.0753	0.0939
ζ	0.0993	0.4880	0.1240	0.3008	0.1386	0.2587
$\zeta(\tau)$	0.2110	0.4645	0.1240	0.3008	0.1386	0.2587
η	0.0108	0.1051	-0.1453	0.1657	-0.1267	0.1560
$\eta(\tau)$	-0.0859	0.0935	-0.1453	0.1657	-0.1267	0.1560
MPE	0.0037	0.0634	-0.0325	0.0676	-0.3239	0.3368
MAPE	0.1537	0.1607	0.1409	0.1517	0.3271	0.3397
coverage	0.0008	0.0014	-0.0044	0.0081	-0.0254	0.0259

Table 3.5: Model comparison: GACQ and GARCH with Student's t(4) errors

The picture is very similar if we consider a DGP with Student errors with four degrees of freedom, multiplied by $1/\sqrt{2}$ in order to normalize the variances. This indicates that for fatter tails the maximum likelihood estimation becomes more difficult for the GARCH model, resulting in increasing prediction errors. Interestingly, the GARCH-N estimator performs somewhat better in terms of predictions then the GARCH-t estimator, even with Student errors. This likely stems from the known overprediction of conditional volatilities in conditional variance models [Klaassen, 2002], which is however compensated by multiplying the latter with the quantile of the Normal distribution instead of the corresponding one of the t(4) distribution, which is larger in magnitude.

Table 3.6 shows simulation results using a Gumbel distribution which belongs to the class of asymmetric distributions and has both positive skewness and excess-kurtosis. The Gumbel distribution is parametrized with location parameter $\mu_G = 0$ and scale parameter $\beta_G = \sqrt{6}/\pi$ which normalizes the variance to one, ensuring comparability of the results. The distribution is then re-centered by subtracting $\beta_g e^1$ from each realization, such that ε has mean zero. Skewness and excess-kurtosis are constant with respect to the distribution's parameters. Given that we are interested in the 95% value at risk and thus estimate $\tau = 0.05$, we mirror the distribution by flipping the signs of the innovations, to evaluate the performance of the longer and heavier tail of the distribution which is then below zero. It can be seen, that the proposed conditional quantile model performs equally well as in most cases with symmetric errors and comparable excess kurtosis. This should come with no surprise, as the quantiles of the error distribution are subject to estimation. It should also be mentioned that the GARCH-N model with two regimes works again surprisingly well in this case, but is outperformed by our conditional quantile estimation procedure.

	GACQ		GARCH-N		GARCH-t	
	Bias	RMSE	Bias	RMSE	Bias	RMSE
β_0^I	0.1362	0.2624	-0.0115	0.0416	-0.0152	0.0393
β_1^I	-0.0287	0.2080	-0.1480	0.2171	-0.1289	0.2224
γ_1^I	-0.0841	0.1920	0.0269	0.1300	0.0088	0.0929
β_0^{II}	0.0614	0.1237	0.0845	0.1996	0.1250	0.1850
β_1^{II}	-0.0360	0.1253	-0.1387	0.1872	-0.1610	0.2076
γ_1^{II}	-0.0032	0.1146	0.0844	0.1766	0.0610	0.1311
ζ	0.2723	0.4382	0.0058	0.6362	0.1175	1.2585
$\zeta(\tau)$	0.1346	0.2927	0.0058	0.6362	0.1175	1.2585
η	0.0575	0.0967	-0.0720	0.1070	-0.0580	0.1053
$\eta(\tau)$	-0.0425	0.0750	-0.0720	0.1070	-0.0580	0.1053
MPE	0.0108	0.0277	0.0419	0.0541	-0.2928	0.3118
MAPE	0.0727	0.0749	0.1213	0.1322	0.2997	0.3172
coverage	0.0007	0.0015	0.0038	0.0090	-0.0268	0.0274

Table 3.6: Model Comparison for re-centered, mirrored Gumbel errors

Next we consider outliers in our innovations. To ensure stability of the series we will introduce them on u_t instead of ε_t in the following way. Let $\mathbf{r} \sim \mathcal{U}[0,1]^n$, then for each $u_t(\theta_0) = \sigma_t(\boldsymbol{z}_t, \theta_0)\varepsilon_t$ the new series $\{u'\}_{t=1}^n$ is defined as

$$u'_t := u_t + \mathbb{1}_{\{r_t \leqslant 0.025\}} \operatorname{sgn}(\varepsilon_t) 3\sigma_\varepsilon$$

with $\sigma_\varepsilon = 1$, due to our normalization. Note that this might be considered a very small contamination, however we report estimates for the 95$^{\text{th}}$ quantile. Thus these contaminated values make up for a large proportion of the data used for estimation.

We only compare our estimator with the correctly specified GARCH estimators for Normal and Student errors.

	GACQ: $\varepsilon \sim \mathcal{N}$		GARCH-N: $\varepsilon \sim \mathcal{N}$		GACQ: $\varepsilon \sim t(4)$		GARCH-t: $\varepsilon \sim t(4)$	
	Bias	RMSE	Bias	RMSE	Bias	RMSE	Bias	RMSE
β_0^I	0.2656	0.6630	-0.3962	0.4538	0.2758	0.7464	-0.0975	0.2817
β_1^I	0.0358	0.4642	0.0647	0.3166	0.1187	0.6819	-0.0614	0.1429
γ_1^I	-0.0749	0.2867	0.0804	0.3435	-0.0615	0.3810	-0.0911	0.2344
β_0^{II}	0.0668	0.2337	0.4326	0.4719	0.1801	0.5841	0.1139	0.2651
β_1^{II}	0.0315	0.2780	-0.0313	0.2028	0.0659	0.5414	-0.0089	0.1635
γ_1^{II}	-0.0010	0.1379	-0.1096	0.1308	-0.0290	0.1701	-0.0689	0.0972
ζ	0.1545	0.5263	1.0059	1.0553	0.1686	0.5461	0.3365	0.5907
$\zeta(\tau)$	0.1832	0.5171	1.0059	1.0553	0.1599	0.5316	0.3365	0.5907
η	0.0678	0.1588	-0.1480	0.3122	0.0262	0.1183	-0.0419	0.5081
$\eta(\tau)$	-0.0445	0.0929	-0.1480	0.3122	-0.0805	0.0939	-0.0419	0.5081
MPE	-0.1023	0.1291	-175.09	1008.6	-0.1528	0.1940	-0.9166	0.9684
MAPE	0.2180	0.2311	175.22	1008.6	0.2749	0.2933	0.9182	0.9696
coverage	0.0007	0.0013	-0.0144	0.0164	0.0009	0.0014	-0.0269	0.0273

Table 3.7: Outliers

It comes with no surprise that the prediction errors somewhat deteriorate compared to the non-outlier case even for the GACQ model. While the GARCH-t specification seems to handle outliers rather well by under-estimating the degrees of freedom parameter (3.0535 with RMSE of 0.97429), the GARCH-N specification breaks down completely in terms of average prediction errors. Note that the coverage ratio in the conditional variance models are on average off by 1.4 and 2.7 percentage points, respectively, which is a rather significant deviation given that we consider the 5% value at risk.

	CQ: (G_{log}, G_{thr})		CH: (G_{log}, G_{thr})		CQ: (G_{log}, G_{lin})		CH: (G_{log}, G_{lin})	
	Bias	RMSE	Bias	RMSE	Bias	RMSE	Bias	RMSE
β_0^I	0.1548	0.3590	0.0375	0.1184	0.1325	0.3504	0.0367	0.1366
β_1^I	0.0153	0.3660	-0.1296	0.1367	0.0325	0.3548	-0.1361	0.1401
γ_1^I	-0.0997	0.2446	-0.1588	0.1938	-0.1000	0.2890	-0.1506	0.1950
β_0^{II}	-0.0415	0.1553	-0.1215	0.1420	-0.0105	0.1587	-0.0870	0.1319
β_1^{II}	0.0661	0.2205	-0.0141	0.1041	0.0335	0.2357	-0.0040	0.1232
γ_1^{II}	0.0241	0.1145	-0.0212	0.1195	0.0164	0.1259	-0.0528	0.1070
ζ	0.1535	0.3663	0.0329	0.2752	0.1757	0.4881	0.0950	0.3580
$\zeta(\tau)$	0.0684	0.2526	0.0329	0.2752	0.0563	0.3104	0.0950	0.3580
η	0.0145	0.1390	-0.1663	0.1712	0.0388	0.1388	-0.0555	0.8775
$\eta(\tau)$	-0.0548	0.0846	-0.1663	0.1712	-0.0523	0.0849	-0.0555	0.8775
MPE	0.0060	0.0528	0.0750	0.0887	0.0143	0.0535	0.0750	0.0887
MAPE	0.1290	0.1336	0.1693	0.1825	0.1201	0.1241	0.1624	0.1757
coverage	0.0006	0.0014	0.0074	0.0107	0.0008	0.0014	0.0074	0.0103

Table 3.8: Misspecified transition functions: (G, G_0) for GACQ and GARCH-N

Since in all our models we assume the transition function to be known *a priori*, in our last experiment we consider the consequences of mis-specifying the model. For this we use two different DGP's, the threshold and the linear transition function. We estimate them using a logistic specification for both GACQ and GARCH-N. The innovation distribution is assumed to be standard normal again.

The elements of the tuple (G, G_0) refer to the transition functions specified for estimation G and the transition function used for the data generating process G_0. It is comforting to report that both models are very robust with respect to misspecified transition functions, at least within the set of parametrized transition functions we consider.

Empirical Application

The following empirical study, in which we again report results for both the proposed ANST-GACQ model and its conditional variance counterparts, is not only supposed to demonstrate how the estimation works with real data, but is also used to evaluate out-of-sample forecasts. The data considered in this application is the GBP/USD exchange rate and the German equity index (DAX) which both are available on a daily basis from 1999 to 2016. The main interest lies in the modelling of the 5% value at risk of the time-series' daily log-returns. We will select a sample of the most recent $n = 3{,}000$

observations each, which corresponds to 12 years, assuming about 250 trading days per year. Model selection is based on a qualitative assessment of the model improvement, when including additional lags, rather than applying formal selection criteria. The specification that is ultimately reported corresponds the widely used specification with $p = q = 1$ lags and a logistic transition function, although estimation turned out to be very robust with respect to the choice of $G \in \mathfrak{G}$. As a transition variable ξ_t we use the lagged dependent variable u_{t-1}. In principal a model selection approach evaluating different legs would be feasible, however since we have financial time series data which is known to react to news very quickly, we have reason to believe that the regime is determined by the previous observation. Parameter estimates of the series are reported in Table 3.9. Recall that these parameters have the structure $\widehat{\theta}_j(\tau) = \widehat{\theta}_j F_{\varepsilon}^{-1}(\tau)$ for $j \in \mathcal{J}_{0,2(p+q+1)}$ and are thus local with respect to the estimated quantile and are in therefore in general negative for lower quantiles. Note however, that the usual parameter restrictions of the GARCH model apply only to the linear combination of both regimes' parameters with weight given by the transition function $G_t(\zeta)$.

With 0.14% the location parameter for USD/GBP is rather close to zero, whereas the one for the German equity index is −0.96%, implying that the dynamic behaviour of the time series changes only after substantial daily declines. These locations correspond to the 68$^{\text{th}}$ and the 17$^{\text{th}}$ unconditional quantile of the respective log-return time series. The scale estimate for the suggests that there is a rapid transition from one regime to the other, around the origin of the transition variable $\xi_t = u_{t-1}$ for both time series.

Regarding the GARCH coefficients for USD/GBP, it should be mentioned that there are relatively large standard errors of $\widehat{\zeta}_n$ the constants in both regimes are not significantly different from zero. However, this does not necessarily indicate that there are not two regimes in this series. In contrast, looking at the estimates for the DAX, we see that the second regime has higher variance than the first one, which indicates that after a turbulent trading day with high losses, we are likely to see a more calm next day driven also by the mean-reverting nature of the first regime, captured by the sign of the $\widehat{\gamma}_n^{\mathrm{I}}$ estimate.

Figure 3.2 shows the last 400 observations of the returns of the German equity index (DAX) and its estimated 5% and 95% conditional quantiles. It can be seen that the symmetry restriction of the conditional variance based ANST-GARCH models, that have been investigated in the previous sections, does not hold.

	USD/GBP						DAX					
	Regime I			Regime II			Regime I			Regime II		
	coef		s.e.	coef		s.e.	coef		s.e.	coef		s.e.
β_0	0.3346	*	0.1742	-0.1540		0.1547	-0.7163	***	0.1543	-1.1583	***	0.1918
β_1	-1.5553	**	0.6695	-1.0954	***	0.3558	-1.3977	***	0.1468	-0.5477	***	0.1772
γ_1	0.3333	**	0.1356	-0.8111	***	0.0656	0.1364	*	0.0785	-0.3192	***	0.0665
ζ	0.1469		0.4404				-0.9636		0.2421			
η	0.1000		0.2799				0.1207		0.2804			

Table 3.9: Coefficients and standard errors for 5%-VaR ANST-GACQ estimates

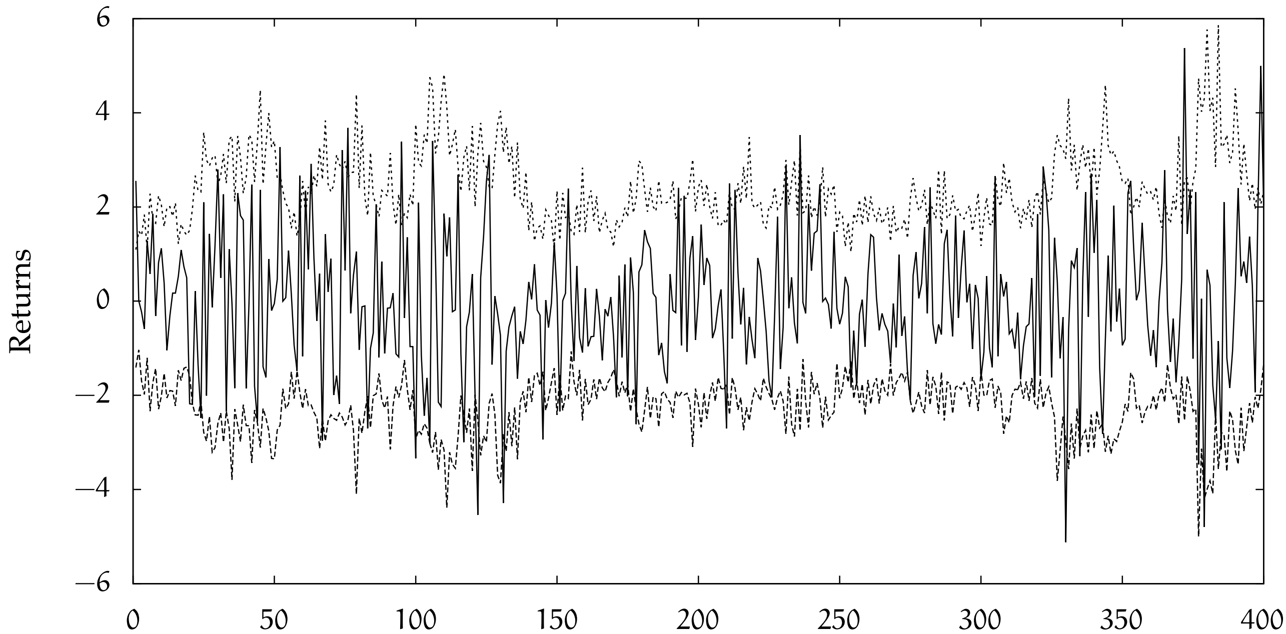

Figure 3.2: DAX returns in % and its corresponding 5% and 95% ANST-GACQ estimates

Since the conditional quantile estimation is an in-sample concept meaning that the coverage ratio holds by construction, in this final step the out-of-sample forecasting is evaluated. For this procedure a subsample consisting of the 1,100 most recent observations of the series is selected. Starting with the estimation of the index-subset $\mathcal{I}_{T_0,T}$ consisting of the first $T - T_0 = 1{,}000$ observations, a forecast $\widehat{Q}_{u_{T+1}}(\tau|\mathcal{F}_T)$ is calculated. Using the true u_{T+1} we calculate a boolean value, which is true if the realization is below the specified quantile. This is repeated for $n = 100$ times by offsetting T_0 by one each time, shifting the window which is considered for estimation. The resulting Bernoulli process $I_{t+1} = \mathbb{1}\left\{u_t \leqslant \widehat{Q}_{u_t}(\tau|\mathcal{F}_{t-1})\right\}$ with $t \in \mathcal{I}_{0,n}$ is the basis for calculating test statistics and the out-of-sample coverage ratio which is easily observed by averaging over t, such that $h_{oos} := n^{-1} \sum_{t=1}^{n} I_t$. As is common in the literature about evaluating value at risk models, two tests are considered to show that the proportion of forecasts exceeding the estimated quantile is not significantly different from τ. We follow the common notion denote such an occurrence as a "hit". First a likelihood ratio test as proposed by Kupiec [1995], which assumes the Bernoulli process to be an i.i.d. sequence, is applied. The probability of having exactly x hits is given by the binomial probability mass function

$$f_{\text{bin}}(x, n, p) = \binom{n}{x} p^x (1-p)^{n-x}$$

Using $x = nh_{\text{oos}}$, the likelihood ratio is given by

$$\begin{aligned}\mathcal{L}_n &= -2\Big[\log f_{\text{bin}}(nh_{\text{oos}}, n, \tau) - \log f_{\text{bin}}(nh_{\text{oos}}, n, h_{\text{oos}})\Big] \\ &= -2n\left[(1-h_{\text{oos}})\log\left(\frac{1-\tau}{1-h_{\text{oos}}}\right) + h_{\text{oos}}\log\left(\frac{\tau}{h_{\text{oos}}}\right)\right]\end{aligned}$$

which is chi-squared distributed with one degree of freedom $\mathcal{L}_n \sim \chi_1^2$ such that the 10%, 5% and 1% critical values are given by 2.7055, 3.8415, and 6.6349, respectively. In addition to this, another formal test is used to support the claim that the out-of-sample forecast does in fact represent the τ^{th} quantile. This test exploits the fact that $I_{t+1} - \tau$ is a martingale difference sequence with zero mean and variance $n\tau(1-\tau)$ and thus its cumulative sum converges to a normal distribution. Formally, the following statistic of the, now two-sided, test can be defined as

$$Z_n := (n\tau(1-\tau))^{\frac{1}{2}} \sum_{i=1}^{n} (I_{t+1} - \tau) \rightsquigarrow \mathcal{N}(0,1) \ \text{ as } n \to \infty$$

with critical values for its absolute value given by 1.6449, 1.9600 and 2.5758, respectively.

The results of this testing procedures are reported in Table 3.10 together with in-sample and out-of-sample coverage ratios. The experiments were repeated for the 5^{th}, 10^{th} and 25^{th} quantile. Looking at the test statistics, it can indeed be concluded, that for our ANST-GACQ model there is no evidence that the null hypothesis, stating that the forecast represents the true quantile, has to be rejected for any of the confidence levels. The situation is different for both standard quadratic ANST-GARCH-N and ANST-GARCH-t models. They both perform poorly for the USD/GBP exchange rate, especially for tail estimates. Although they improve as more centered quantiles are considered, the estimates are significantly different from the true values estimating the 10^{th} quantile in the ANST-GARCH-N model and for the 5^{th}, 10^{th} and 25^{th} quantile in the ANST-GARCH-t model, considering a 90% confidence interval. On the other hand, log-returns of the DAX equity index seem to be well-represented by a ANST-

USD/GBP Exchange rate

	GACQ			GARCH-N			GARCH-t		
	5%	10%	25%	5%	10%	25%	5%	10%	25%
h_{IS}	0.0520	0.1009	0.2488	0.0228	0.0337	0.1292	0.0118	0.0282	0.1119
h_{OOS}	0.0300	0.0700	0.2500	0.0200	0.0500	0.2100	0.0100	0.0400	0.1800
Z_n	-0.9177	-1.0000	0.0000	-1.3765	-1.6667*	-0.9238	-1.8353*	-2.0000*	-1.6161
$\mathcal{L}_n$	0.9769	1.1055	0.0000	2.4286	3.3413*	0.8868	4.9472**	5.0611**	2.8078*

DAX Equity Index

	CQ			GARCH-N			GARCH-t		
	5%	10%	25%	5%	10%	25%	5%	10%	25%
h_{IS}	0.0500	0.0999	0.2498	0.1410	0.1747	0.2767	0.1383	0.1705	0.2757
h_{OOS}	0.0600	0.1000	0.2500	0.0500	0.1200	0.2300	0.0800	0.1500	0.2400
Z_n	0.4589	0.0000	0.0000	0.0000	0.6667	-0.4619	1.3765	1.6667*	-0.2309
$\mathcal{L}_n$	0.1984	0.0000	0.0000	0.0000	0.4205	0.2173	1.6158	2.4470	0.0538

Table 3.10: Coverage and Test statistics: Comparison to GARCH

GARCH-N(1,1) process as well, as it yields rather precise forecasts for all considered quantiles. From these very mixed results about the performance of the conditional variance models it can be concluded, that each of the considered time series exhibits a distinct dynamic behaviour. This provides a very strong argument for the ANST-GACQ model, as it is robust with respect to these individual time series characteristics and performs equally well in both cases.

To actually compare our ANST-GACQ model with the single regime GACQ model we report separate out-of-sample forecasts for the latter in Table 3.11, instead of developing a formal test for the regime-switching model against the linear model (see e.g. Luukkonen et al. [1988] for the smooth transition autogregressive model).

	USD/GBP				DAX Equity Index			
	h_{IS}	h_{OOS}	Z_n	$\mathcal{L}_n$	h_{IS}	h_{OOS}	Z_n	$\mathcal{L}_n$
25%	0.24875	0.24	-0.2309	0.0538	0.24880	0.26	0.2309	0.0529
10%	0.09890	0.05	-1.6667 *	3.3413 *	0.09890	0.14	1.3333	1.6017
5%	0.04995	0.01	-1.8353 *	4.9472 **	0.04995	0.06	0.4588	0.1984

Table 3.11: Coverage and Test statistics: Comparison to single regime

One can see that in this experiment with only one regime, we find that for the USD/GBP time series we would reject the hypothesis that the proportion of hits equals the specified value at risk in our out-of-sample forecast procedure, at least for quantiles further out in the tail. This was not the case for the two-regime model, which is a strong indication that there is indeed a second regime in this time-series. We do not find this for the German equity index. Nevertheless, from this we cannot necessarily conclude that there might not be a second regime present also in this time series. It merely means that out-of-sample forecasting works satisfactorily if only regime is considered, although the two-regime model provides slightly better forecasts, especially at the 10% level. Considering the location we find in our two-regime model, we found that there are only 17% of the observation in the first regime. Thus the time series is mostly driven by the second regime which might be the reason for the relatively good performance of its one-regime counterpart.

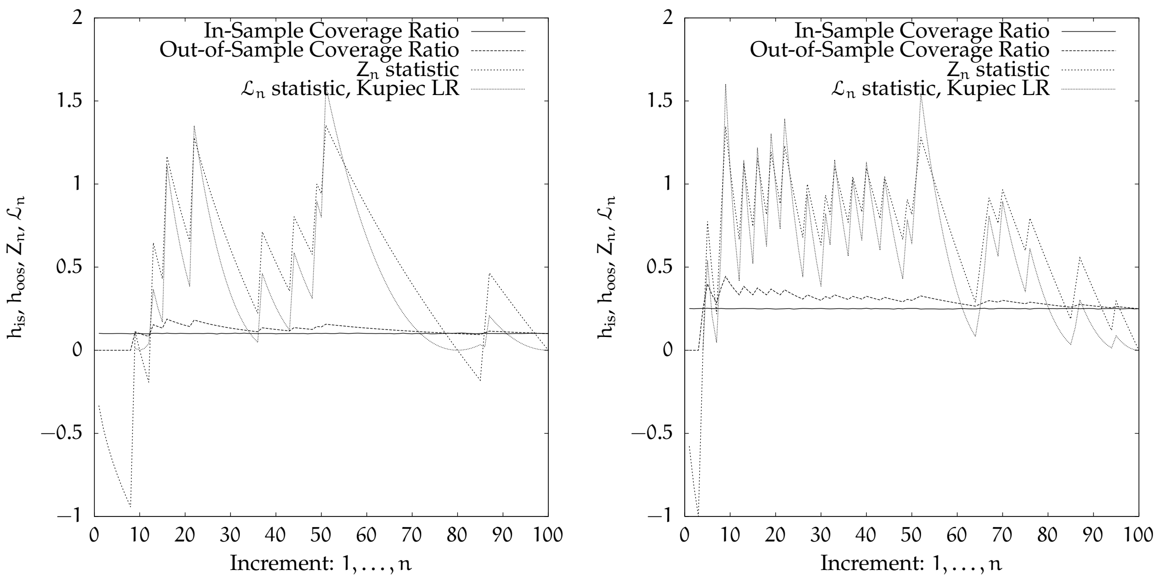

Figure 3.3: DAX: Convergence for Coverage Tests, 5% (l.h.s) and 25% (r.h.s.)

Figure 3.3 shows for the two-regime specification the convergence of the coverage ratio and the corresponding test statistics for the 5^{th} and 25^{th} quantile of USD/GBP, respectively, to demonstrate how fast these statistics actually converge to a value close to zero.

Conclusion and Outlook

In the preceding analysis a time-series model of conditional quantiles has been proposed allowing for two parameter regimes with smooth transition between them. Compared to asymmetric non-linear smooth transition GARCH models, in which conditional variance is estimated to calculate conditional quantiles, by imposing a distributional restriction, modelling conditional quantiles directly has the advantage of not relying on such an assumption. Furthermore, it has been shown that the estimation procedure has the advantage of being robust with respect to large innovations. In contrast to the one-regime conditional quantile GARCH model and even the asymmetric CAViaR specification, the proposed generalized asymmetric non-linear smooth transition model allows for additional flexibility especially with respect to asymmetric dynamic responses to past quantiles. In addition to this, it has been shown that out-of-sample forecasts are very accurate and superior to the ones of both the one-regime case and the smooth transition conditional variance model, considering common proportion of failure tests. However, the estimation procedure does come with some limitations. While it is possible to use any function of past observations as transition variable, the latter is only allowed to have a one-dimensional range. Higher dimensional transition variables would be feasible and easy to incorporate from a theoretical point of view, however from a computational point of view this would be very expensive, due to the exhaustive search algorithm used to identify them. A possible extension of the model would be to allow the transition variable to be a function of past scales or quantiles, as it is done for conditional variance in Lanne [2005].

Auxiliary Results and Proofs

Within the following derivations, let $\mathbb{P}_n$ be the empirical distribution that puts mass $d\mathbb{P}_n = n^{-1}$ to each observation $u_1, \dots, u_n$ such that $\mathbb{P}_n f(u_t) = \int f(u_t) d\mathbb{P}_n = n^{-1} \sum_{t=1}^{n} f(u_t)$ for any measurable function f. Also let the vector $z_t^m = [|u_{t-1}|, \dots, |u_{t-m}|]^T$, and $z_t(\zeta)^T = [G_t(\zeta)\, z_t^m, (1 - G_t(\zeta)) z_t^m]$, where the latter is defined as a function of ζ to highlight the dependence on the transition function. Further let the data considered in the second stage be $z_t^{pq} = z_t(\alpha) = [1, |u_{t-1}|, \dots, |u_{t-p}|, \sigma_{t-1}(\alpha), \dots, \sigma_{t-q}(\alpha)]^T$ which is a function of $\alpha^T := [\alpha^{I,T}, \alpha^{II,T}, \zeta^T]^T$ where ζ refers to the location and scale parameters entering the first stage. Similarly, we define $z_t(\alpha, \zeta)^T = [G_t(\zeta)\, z_t(\alpha), (1 - G_t(\zeta)) z_t(\alpha)]$, to be the vector that stacks both regimes weighted data. Further let $\mathbf{a}^T = [\alpha^{I,T}, \alpha^{II,T}, q^T, \zeta^T]^T$. The right-side derivative of the check-function $\rho_{\tau_k}(u) = u(\tau_k - \mathbb{1}\{u \leqslant 0\})$ is given by $\psi_{\tau_k,t}(u) := (\tau_k - \mathbb{1}\{u \leqslant 0\})$ such that the directional derivative of the objective function defined in equation (3.9) is given as

$$
g_n(\alpha^I, \alpha^{II}, q, \zeta) := \mathbb{P}_n \sum_{k=1}^{K} \underbrace{\begin{bmatrix} z_t(\zeta) q_k \\ \mathbb{1}\{k = 1\}\, [\alpha^I, \alpha^{II}]^T z_t(\zeta) \\ \vdots \\ \mathbb{1}\{k = K\}\, [\alpha^I, \alpha^{II}]^T z_t(\zeta) \\ q_k(\alpha^I - \alpha^{II})^T z_t(\zeta) \frac{\partial G_t(\zeta)}{\partial \zeta} \end{bmatrix}}_{x_{t,k}(\mathbf{a})} (\tau_k - \mathbb{1}\left\{ u_t \leqslant q_k [\alpha^I, \alpha^{II}]^T z_t(\zeta) \right\}). \tag{3.14}
$$

Similarly, without making any statements about convergence yet, the corresponding population equivalent can be written as

$$g(\boldsymbol{\alpha}^{\mathrm{I}}, \boldsymbol{\alpha}^{\mathrm{II}}, \mathbf{q}, \zeta) := \mathbb{E} \sum_{k=1}^{K} \begin{bmatrix} z_t(\zeta) q_k \\ \mathbb{1}\{k=1\} [\boldsymbol{\alpha}^{\mathrm{I}}, \boldsymbol{\alpha}^{\mathrm{II}}]^{\mathsf{T}} z_t(\zeta) \\ \vdots \\ \mathbb{1}\{k=K\} [\boldsymbol{\alpha}^{\mathrm{I}}, \boldsymbol{\alpha}^{\mathrm{II}}]^{\mathsf{T}} z_t(\zeta) \\ q_k (\boldsymbol{\alpha}^{\mathrm{I}} - \boldsymbol{\alpha}^{\mathrm{II}})^{\mathsf{T}} z_t(\zeta) \frac{\partial G_t(\zeta)}{\partial \zeta} \end{bmatrix} (\tau_k - F_{u_t|\mathcal{F}_{t-1}}(q_k [\boldsymbol{\alpha}^{\mathrm{I}}, \boldsymbol{\alpha}^{\mathrm{II}}]^{\mathsf{T}} z_t(\zeta)))$$

by applying the law of iterated expectations. In addition to this we will frequently make use of the identity $\partial F_{u_t|\mathcal{F}_{t-1}}(x)/\partial x = {\sigma_t}^{-1} \partial F_\varepsilon(x)/\partial x = \sigma_t^{-1} f_\varepsilon(x)$. Equation (3.5) establishes $F^{-1}_{u_t|\mathcal{F}_{t-1}}(x) = \sigma_t F_\varepsilon^{-1}(x)$ and since both $F_{u_t|\mathcal{F}_{t-1}}$ and F_ε are monotone and differentiable by Assumption 3.3 and 3.4, the expression follows by applying the inverse function theorem.

Theorem 3.1 (Identification and First Stage Consistency)**.** Let $\mathbf{a}^{\mathsf{T}} = [\boldsymbol{\alpha}^{\mathrm{I},\mathsf{T}}, \boldsymbol{\alpha}^{\mathrm{II},\mathsf{T}}, \mathbf{q}^{\mathsf{T}}, \zeta^{\mathsf{T}}]^{\mathsf{T}}$. Under Assumptions 3.1-3.8, it holds for $n \to \infty$

$$\|\widehat{\mathbf{a}}_n - \mathbf{a}_0\| = \mathcal{O}_p\left(\frac{m}{n}\right).$$

Proof. We split up the identification and consistency argument into two parts. First we show that for $\boldsymbol{\alpha}(\tau) = [\boldsymbol{\alpha}^{\mathrm{I},\mathsf{T}}(\tau), \boldsymbol{\alpha}^{\mathrm{II},\mathsf{T}}(\tau)]$, the vector $\left[\boldsymbol{\alpha}_0^{\mathsf{T}}(\tau), \zeta_0^{\mathsf{T}}\right]^{\mathsf{T}}$ is in fact the minimum of the objective function $\mathbb{E}\rho_\tau(u)$. We show this for an arbitrary quantile $\tau \in (0, \frac{1}{2}) \cup (\frac{1}{2}, 1)$, from which the composite quantile result follows.

To see this let $m(\boldsymbol{\alpha}(\tau), \zeta) = \mathbb{E}\rho_\tau(u_t - \boldsymbol{\alpha}^{\mathsf{T}}(\tau) z_t(\zeta))$. Then we have global identification if and only if $m(\boldsymbol{\alpha}(\tau), \zeta) - m(\boldsymbol{\alpha}_0(\tau), \zeta_0) > 0$ for any $\left[\boldsymbol{\alpha}^{\mathsf{T}}(\tau), \zeta^{\mathsf{T}}\right]^{\mathsf{T}} \neq \left[\boldsymbol{\alpha}_0^{\mathsf{T}}(\tau), \zeta_0^{\mathsf{T}}\right]^{\mathsf{T}}$.

Let $\boldsymbol{\alpha}(\tau)^{\Delta} = \boldsymbol{\alpha}^{\mathrm{I}}(\tau) - \boldsymbol{\alpha}^{\mathrm{II}}(\tau)$ and also let ν_t be the probability measure of u_t conditional upon the filtration $\mathcal{F}_{t-1}$. Then we have to prove

$$\inf_{\substack{\boldsymbol{\alpha}(\tau) : \|\boldsymbol{\alpha}(\tau) - \boldsymbol{\alpha}_0(\tau)\| > \delta \\ \zeta : \|\zeta - \zeta_0\| > \delta}} \mathbb{E}\left[\int \rho_\tau(u_t - \boldsymbol{\alpha}(\tau)^{\mathsf{T}} z_t(\zeta)) d\nu_t - \int \rho_\tau(u_t - \boldsymbol{\alpha}_0(\tau)^{\mathsf{T}} z_t(\zeta_0)) d\nu_t\right] > \varepsilon_\delta$$

where we can split up the part inside the expectation into two cases. First consider $\boldsymbol{\alpha}_0(\tau)^{\mathsf{T}} z_t(\zeta_0) >$

$\alpha(\tau)^\mathsf{T} z_t(\zeta)$ such that

$$\int \rho_\tau(u_t - \alpha(\tau)^\mathsf{T} z_t(\zeta))d\nu_t - \int \rho_\tau(u_t - \alpha_0(\tau)^\mathsf{T} z_t(\zeta_0))d\nu_t =$$

$$(\tau-1)\int_{-\infty}^{\alpha(\tau)^\mathsf{T} z_t(\zeta)} (u_t - \alpha(\tau)^\mathsf{T} z_t(\zeta))d\nu_t + \tau \int_{\alpha(\tau)^\mathsf{T} z_t(\zeta)}^{\alpha_0(\tau)^\mathsf{T} z_t(\zeta_0)} (u_t - \alpha(\tau)^\mathsf{T} z_t(\zeta))d\nu_t + \tau \int_{\alpha_0(\tau)^\mathsf{T} z_t(\zeta_0)}^{+\infty} (u_t - \alpha(\tau)^\mathsf{T} z_t(\zeta))d\nu_t$$
$$-(\tau-1)\int_{-\infty}^{\alpha(\tau)^\mathsf{T} z_t(\zeta)} (u_t - \alpha_0(\tau)^\mathsf{T} z_t(\zeta_0))d\nu_t - \tau \int_{\alpha(\tau)^\mathsf{T} z_t(\zeta)}^{\alpha_0(\tau)^\mathsf{T} z_t(\zeta_0)} (u_t - \alpha_0(\tau)^\mathsf{T} z_t(\zeta_0))d\nu_t - \tau \int_{\alpha_0(\tau)^\mathsf{T} z_t(\zeta_0)}^{+\infty} (u_t - \alpha_0(\tau)^\mathsf{T} z_t(\zeta_0))d\nu_t$$

$$=(1-\tau)\int_{-\infty}^{\alpha(\tau)^\mathsf{T} z_t(\zeta)} (\alpha(\tau)^\mathsf{T} z_t(\zeta) - \alpha_0(\tau)^\mathsf{T} z_t(\zeta_0))d\nu_t + \Omega_1 + \tau \int_{\alpha_0(\tau)^\mathsf{T} z_t(\zeta_0)}^{+\infty} (\alpha_0(\tau)^\mathsf{T} z_t(\zeta_0) - \alpha(\tau)^\mathsf{T} z_t(\zeta))d\nu_t$$
$$\geqslant(1-\tau)\int_{-\infty}^{\alpha_0(\tau)^\mathsf{T} z_t(\zeta_0)} (\alpha(\tau)^\mathsf{T} z_t(\zeta) - \alpha_0(\tau)^\mathsf{T} z_t(\zeta_0))d\nu_t + \tau \int_{\alpha_0(\tau)^\mathsf{T} z_t(\zeta_0)}^{+\infty} (\alpha_0(\tau)^\mathsf{T} z_t(\zeta_0) - \alpha(\tau)^\mathsf{T} z_t(\zeta))d\nu_t \quad (3.15)$$
$$=\alpha_0(\tau)^\mathsf{T} z_t(\zeta_0) - \alpha(\tau)^\mathsf{T} z_t(\zeta)\left[(1-\tau)\int_{-\infty}^{\alpha_0(\tau)^\mathsf{T} z_t(\zeta_0)} d\nu_t + \tau \int_{\alpha_0(\tau)^\mathsf{T} z_t(\zeta_0)}^{+\infty} d\nu_t\right]$$
$$= \left(\alpha_0(\tau)^\mathsf{T} z_t(\zeta_0) - \alpha(\tau)^\mathsf{T} z_t(\zeta)\right)\left[\tau - \nu_t(-\infty, F^{-1}_{u_t|\mathcal{F}_{t-1}}(\tau)]\right], \quad (3.16)$$

where we use the fact that $\nu_t(-\infty, F^{-1}_{u_t|\mathcal{F}_{t-1}}(\tau)] + \nu_t(F^{-1}_{u_t|\mathcal{F}_{t-1}}(\tau), +\infty) = 1$, since ν_t is a probability measure, and where the inequality in equation (3.15) follows from the fact that

$$\Omega_1 := \int_{\alpha(\tau)^\mathsf{T} z_t(\zeta)}^{\alpha_0(\tau)^\mathsf{T} z_t(\zeta_0)} (u_t - \tau\alpha(\tau)^\mathsf{T} z_t(\zeta) - (1-\tau)\alpha_0(\tau)^\mathsf{T} z_t(\zeta_0))d\nu_t \geqslant (1-\tau)\int_{\alpha(\tau)^\mathsf{T} z_t(\zeta)}^{\alpha_0(\tau)^\mathsf{T} z_t(\zeta_0)} (\alpha(\tau)^\mathsf{T} z_t(\zeta) - \alpha_0(\tau)^\mathsf{T} z_t(\zeta_0))d\nu_t.$$

Similarly for $\alpha_0(\tau)^\mathsf{T} z_t(\zeta_0) < \alpha(\tau)^\mathsf{T} z_t(\zeta)$ it holds

$$=(1-\tau)\int_{-\infty}^{\alpha_0(\tau)^\mathsf{T} z_t(\zeta_0)} (\alpha(\tau)^\mathsf{T} z_t(\zeta) - \alpha_0(\tau)^\mathsf{T} z_t(\zeta_0))d\nu_t - \Omega_2 + \tau \int_{\alpha(\tau)^\mathsf{T} z_t(\zeta)}^{+\infty} (\alpha_0(\tau)^\mathsf{T} z_t(\zeta_0) - \alpha(\tau)^\mathsf{T} z_t(\zeta))d\nu_t$$
$$\geqslant(1-\tau)\int_{-\infty}^{\alpha_0(\tau)^\mathsf{T} z_t(\zeta_0)} (z_t(\zeta)\alpha(\tau)^\mathsf{T} - z_t(\zeta_0)\alpha_0(\tau)^\mathsf{T})d\nu_t + \tau \int_{\alpha_0(\tau)^\mathsf{T} z_t(\zeta_0)}^{+\infty} (\alpha_0(\tau)^\mathsf{T} z_t(\zeta_0) - \alpha(\tau)^\mathsf{T} z_t(\zeta))d\nu_t$$
$$= \left(\alpha(\tau)^\mathsf{T} z_t(\zeta) - \alpha_0(\tau)^\mathsf{T} z_t(\zeta_0)\right)\left[\tau - \nu_t(-\infty, F^{-1}_{u_t|\mathcal{F}_{t-1}}(\tau)]\right], \quad (3.17)$$

where we use

$$\Omega_2 := \int_{\alpha_0(\tau)^\mathsf{T} z_t(\zeta_0)}^{\alpha(\tau)^\mathsf{T} z_t(\zeta)} (u_t - \tau\alpha(\tau)^\mathsf{T} z_t(\zeta) - (1-\tau)\alpha_0(\tau)^\mathsf{T} z_t(\zeta_0))d\nu_t \leqslant \tau \int_{\alpha_0(\tau)^\mathsf{T} z_t(\zeta_0)}^{\alpha(\tau)^\mathsf{T} z_t(\zeta)} (\alpha_0(\tau)^\mathsf{T} z_t(\zeta_0) - \alpha(\tau)^\mathsf{T} z_t(\zeta))d\nu_t.$$

Then, by the definition of the τ^{th} quantile we have have that $\nu_t(-\infty, F^{-1}_{u_t|\mathcal{F}_{t-1}}(\tau)] \leqslant \tau$ such that the final expressions in both equation (3.16) and equation (3.17) are non-negative. Thus the expectation with respect to the measure of z_t is zero for parameters other than the true parameter if and only if for all z_t we have that $\left(\alpha_0(\tau)^T z_t(\zeta_0) - \alpha(\tau)^T z_t(\zeta)\right) = 0$ from which we get our identification statement:

$$\begin{aligned}\mathbb{E}\left[\alpha(\tau)^T z_t(\zeta) - \alpha_0(\tau)^T z_t(\zeta_0)\right] &= \mathbb{E}\left[(\alpha-\alpha_0)^T F_\varepsilon^{-1}(\tau) z_t(\zeta) + \alpha_0^\Delta F_\varepsilon^{-1}(\tau) z_t^m (G_t(\zeta) - G_t(\zeta_0))\right] \\ &= F_\varepsilon^{-1}(\tau)\mathbb{E}\left[\begin{pmatrix}\alpha-\alpha_0 \\ \alpha^\Delta\end{pmatrix}^T \begin{pmatrix} z_t(\zeta) \\ z_t^m(G_t(\zeta)-G_t(\zeta_0))\end{pmatrix}\right] = 0\end{aligned}$$

if and only if $\alpha = \alpha_0$ and $\zeta = \zeta_0$. To ensure this we require the global identification assumption stating that $\mathbb{E}[G_t(\zeta)\, z_t^m, (1-G_t(\zeta))z_t^m, z_t^m(G_t(\zeta)-G_t(\zeta_0))]^T[G_t(\zeta)\, z_t^m, (1-G_t(\zeta))z_t^m, z_t^m(G_t(\zeta)-G_t(\zeta_0))]$ has full rank for any ζ which holds by Assumption 3.6. Thus $\alpha(\tau) = \alpha F_\varepsilon^{-1}(\tau)$ is identified for an arbitrary $\tau \in (0, \frac{1}{2}) \cup (\frac{1}{2}, 1)$. Note that the result does not hold for $\tau = \frac{1}{2}$ for which $F_\varepsilon^{-1}(\tau)$ is zero.

Given this identification result, for consistency we need to show that the sample objective function, which is continuous, converges uniformly in probability to the population counterpart, which we discussed above [Newey & McFadden, 1994, Theorem 2.1, p. 2121]. For this note that for $\tau \in (0, \frac{1}{2}) \cup (\frac{1}{2}, 1)$ the parameter space Θ_1^τ is a compact subset of $\mathbb{R}^{2(m+1)+2}$ and denote $m = \rho_\tau \circ \omega_t$, where we let $\omega_t : (\alpha^I(\tau), \alpha^{II}(\tau), \zeta) \mapsto z_t^{m,T}\alpha^{II}(\tau) + (\alpha^I(\tau) - \alpha^{II}(\tau))^T z_t^m G(\xi_t, \zeta, \eta)$, $\alpha(\tau) = [\alpha^I(\tau, T), \alpha^{II,T}(\tau)]^T$ and $\zeta = [\zeta, \eta]^T$. For consistency we can apply Theorem 1 in Chen & Shen [1998, p. 297], for which we have to check conditions A.1-A.4 therein. First, for condition A.4 we show that the function $m = \rho_\tau \circ \omega_t$ is Lipschitz. By Assumption 3.1 the transition function $G_t(\zeta)$ is Lipschitz in ζ and by construction so is ω_t in α and $G()$. Thus ω_t is Lipschitz in α and ζ since the property is preserved under function composition[5]. The piece-wise linear function ρ_τ is Lipschitz as well, and hence so is $m = \rho_\tau \circ \omega_t$. Further, note that we actually have a special case of $s = 1$ in the Hölder condition A.4, which by Chen & Shen [1998, Remark 1(c)] implies their condition A.2. In addition to this, condition A.1 holds by Assumption 3.5 which states that u_t is β-mixing with decay rate satisfying $\beta_s \leqslant \beta_0 s^{-(2+\delta)}$ for $\delta > 0$. Finally, we note that by Chen [2008, p. 5595] it holds for Lipschitz functions that $\log N_{[]}(\epsilon^s, \mathcal{F}_n, \|\cdot\|_2) \leqslant \log N(\epsilon, \Theta_1^\tau, \|\cdot\|) \leqslant Cm\log(\frac{1}{\epsilon})$ for some constant $C > 0$, where m is the dimension of the sieve parameter space, which implies their condition A.3. Thus, we can apply Theorem 1 in Chen & Shen [1998] and conclude that our first stage sieve estimator is consistent.

Next, we have to show that the directional derivative of the non-differentiable g_n around the true value a_0 is positive in every direction with probability tending to one. Let $\varphi_t(a) = \sum_{k=1}^{K} x_t(a)(\tau_k - \mathbb{1}\left\{u_t \leqslant a^T x_t(a)\right\})$. Then we have to show that

$$\forall \epsilon > 0 : \exists B < \infty : \lim_{n\to\infty} \mathbf{P}\left[\inf_{\lambda \in \mathbb{R}^{2(m+1)+K+2}:\|\lambda\|=1} \mathbb{P}_n \lambda^T \varphi_t\left(a_0 + B\left(\frac{m}{n}\right)^{\frac{1}{2}}\lambda\right) > 0\right] > 1-\epsilon. \tag{3.18}$$

Adding and subtracting $\mathbb{P}_n\lambda^T\mathbb{E}\left[\varphi_t\left(a_0 + B\left(\frac{m}{n}\right)^{\frac{1}{2}}\lambda\right)\Big|\mathcal{F}_{t-1}\right] - \mathbb{P}_n\lambda^T\mathbb{E}\left[\varphi_t(a_0)|\mathcal{F}_{t-1}\right]$ as well as $\mathbb{P}_n\lambda^T\varphi_t(a_0)$,

[5]If both f and g are Lipschitz so is $f\circ g$ since $(f\circ g)(x) - (f\circ g)(x_0) = f(g(x)) - f(g(x_0)) \leqslant C_f(g(x) - g(x_0)) \leqslant C_f C_g(x - x_0)$ for finite constants C_f and C_g.

the random quantity in (3.18) can be rewritten as:

$$
\begin{aligned}
&\mathbb{P}_n \boldsymbol{\lambda}^\mathsf{T} \varphi_t \left(\boldsymbol{a}_0 + B \left(\frac{m}{n} \right)^{\frac{1}{2}} \boldsymbol{\lambda} \right) && (3.19)\\
&\quad = \left\{ \mathbb{P}_n \boldsymbol{\lambda}^\mathsf{T} \varphi_t \left(\boldsymbol{a}_0 \right) \right\} && (3.20)\\
&\quad + \left\{ \left(\mathbb{P}_n \boldsymbol{\lambda}^\mathsf{T} \varphi_t \left(\boldsymbol{a}_0 + B \left(\frac{m}{n} \right)^{\frac{1}{2}} \boldsymbol{\lambda} \right) - \mathbb{P}_n \boldsymbol{\lambda}^\mathsf{T} \varphi_t \left(\boldsymbol{a}_0 \right) \right) \right. && (3.21)\\
&\quad - \left. \left(\mathbb{P}_n \boldsymbol{\lambda}^\mathsf{T} \mathbb{E} \left[\varphi_t \left(\boldsymbol{a}_0 + B \left(\frac{m}{n} \right)^{\frac{1}{2}} \boldsymbol{\lambda} \right) \middle| \mathcal{F}_{t-1} \right] - \mathbb{P}_n \boldsymbol{\lambda}^\mathsf{T} \mathbb{E} \left[\varphi_t \left(\boldsymbol{a}_0 \right) \middle| \mathcal{F}_{t-1} \right] \right) \right\} \\
&\quad + \left\{ \mathbb{P}_n \boldsymbol{\lambda}^\mathsf{T} \mathbb{E} \left[\varphi_t \left(\boldsymbol{a}_0 + B \left(\frac{m}{n} \right)^{\frac{1}{2}} \boldsymbol{\lambda} \right) \middle| \mathcal{F}_{t-1} \right] - \mathbb{P}_n \boldsymbol{\lambda}^\mathsf{T} \mathbb{E} \left[\varphi_t \left(\boldsymbol{a}_0 \right) \middle| \mathcal{F}_{t-1} \right] \right\}. && (3.22)
\end{aligned}
$$

This can now be analyzed term by term. Term (3.21) will turn out to be stochastically negligible whereas term (3.20) and (3.22) can be made explicit. This can be seen by writing (3.22) in the following way, using the definition of φ_t,

$$
\begin{aligned}
&\mathbb{P}_n \boldsymbol{\lambda}^\mathsf{T} \mathbb{E} \left[\varphi_t \left(\boldsymbol{a}_0 + B \left(\frac{m}{n} \right)^{\frac{1}{2}} \boldsymbol{\lambda} \right) \middle| \mathcal{F}_{t-1} \right] - \mathbb{P}_n \boldsymbol{\lambda}^\mathsf{T} \mathbb{E} \left[\varphi_t \left(\boldsymbol{a}_0 \right) \middle| \mathcal{F}_{t-1} \right] \\
&= \mathbb{P}_n \boldsymbol{\lambda}^\mathsf{T} \mathbb{E} \left[\sum_{k=1}^{K} \left(\tau_k - \mathbb{1} \left\{ u_t \leqslant \left(\boldsymbol{a}_0 + B \left(\frac{m}{n} \right)^{\frac{1}{2}} \boldsymbol{\lambda} \right)^\mathsf{T} \boldsymbol{x}_{t,k}(\boldsymbol{a}_0) \right\} \boldsymbol{x}_{t,k}(\boldsymbol{a}_0) \right) \middle| \mathcal{F}_{t-1} \right] \\
&- \mathbb{P}_n \boldsymbol{\lambda}^\mathsf{T} \mathbb{E} \left[\sum_{k=1}^{K} \left(\tau_k - \mathbb{1} \left\{ u_t \leqslant \boldsymbol{a}_0^\mathsf{T} \boldsymbol{x}_{t,k}(\boldsymbol{a}_0) \right\} \boldsymbol{x}_{t,k}(\boldsymbol{a}_0) \right) \middle| \mathcal{F}_{t-1} \right] \\
&= -\mathbb{P}_n \boldsymbol{\lambda}^\mathsf{T} \sum_{k=1}^{K} \left(F_{u_t|\mathcal{F}_{t-1}} \left(\left(\boldsymbol{a}_0 + B \left(\frac{m}{n} \right)^{\frac{1}{2}} \boldsymbol{\lambda} \right)^\mathsf{T} \boldsymbol{x}_{t,k}(\boldsymbol{a}_0) \right) - F_{u_t|\mathcal{F}_{t-1}} \left(\boldsymbol{a}_0^\mathsf{T} \boldsymbol{x}_{t,k}(\boldsymbol{a}_0) \right) \right) \boldsymbol{x}_{t,k}(\boldsymbol{a}_0).
\end{aligned}
$$

Applying the Taylor-series expansion around $\boldsymbol{a}_0$ to the first term, for each $t \in \mathcal{I}_{m,n}$ yields:

$$
\begin{aligned}
&- \mathbb{P}_n \boldsymbol{\lambda}^\mathsf{T} \sum_{k=1}^{K} f_{u_t|\mathcal{F}_{t-1}} \left(\boldsymbol{a}_0^\mathsf{T} \boldsymbol{x}_{t,k}(\boldsymbol{a}_0) \right) \boldsymbol{x}_{t,k}(\boldsymbol{a}_0) \boldsymbol{x}_{t,k}(\boldsymbol{a}_0)^\mathsf{T} B \left(\frac{m}{n} \right)^{\frac{1}{2}} \boldsymbol{\lambda} + \mathcal{O}_p \left(m^2 n^{-2} \right) \\
&= -B \left(\frac{m}{n} \right)^{\frac{1}{2}} \boldsymbol{\lambda}^\mathsf{T} \mathbb{P}_n \left[\sum_{k=1}^{K} \frac{1}{\sigma_t} f_\varepsilon \left(F_\varepsilon^{-1}(\tau_k) \right) \boldsymbol{x}_{t,k}(\boldsymbol{a}_0) \boldsymbol{x}_{t,k}(\boldsymbol{a}_0)^\mathsf{T} \right] \boldsymbol{\lambda} + \mathcal{O}_p \left(m^2 n^{-2} \right) \\
&= \left(\frac{m}{n} \right)^{\frac{1}{2}} \boldsymbol{\lambda}^\mathsf{T} \mathbf{D}_{1,n,m} \boldsymbol{\lambda} + \mathcal{O}_p \left(m^2 n^{-2} \right),
\end{aligned}
$$

using equations (3.7), (3.5) for the last step and defining $\mathbf{D}_{1,n,m} :=$

$$
-\mathbb{E}\mathbb{P}_n \frac{1}{\sigma_t} \left[\begin{array}{c:c:c}
\boldsymbol{\iota}_K^\mathsf{T} (\mathbf{s} \odot \mathbf{q} \odot \mathbf{q}) \nu(\zeta) \nu(\zeta)^\mathsf{T} \otimes \boldsymbol{z}_t \boldsymbol{z}_t^\mathsf{T} & \boldsymbol{z}_t^\mathsf{T} \bar{\alpha}(\zeta) (\nu(\zeta) \otimes \boldsymbol{z}_t) \otimes (\mathbf{s} \odot \mathbf{q})^\mathsf{T} & \boldsymbol{\iota}_K^\mathsf{T} (\mathbf{s} \odot \mathbf{q} \odot \mathbf{q}) \boldsymbol{\alpha}^\Delta \boldsymbol{z}_t (\nu(\zeta) \otimes \boldsymbol{z}_t) \frac{\partial G}{\partial \zeta^\mathsf{T}} \\ \hdashline
 & \operatorname{diag}(\mathbf{s}) (\bar{\boldsymbol{\alpha}}^\mathsf{T} \boldsymbol{z}_t)^2 & \bar{\boldsymbol{\alpha}}^\mathsf{T} \boldsymbol{z}_t(\zeta) \boldsymbol{\alpha}^\Delta \boldsymbol{z}_t (\mathbf{s} \odot \mathbf{q}) \frac{\partial G}{\partial \zeta^\mathsf{T}} \\ \hdashline
 & & \boldsymbol{\iota}_K^\mathsf{T} (\mathbf{s} \odot \mathbf{q} \odot \mathbf{q}) (\boldsymbol{\alpha}^\Delta \boldsymbol{z}_t)^2 \frac{\partial G}{\partial \zeta} \frac{\partial G}{\partial \zeta^\mathsf{T}}
\end{array} \right]
$$

where $\nu(\zeta) = [G_t(\zeta), 1 - G_t(\zeta)]^\mathsf{T}$ and $\odot$ is the Hadamard product.

To analyze the remaining terms, let $\eta_t(\nu) = \varphi_t(\boldsymbol{a}_0 + \boldsymbol{\nu}) - \varphi_t(\boldsymbol{a}_0)$. Then term (3.21) is negligible in probability with rate $\left(\frac{m}{n}\right)^{-\frac{1}{2}}$ if

$$
\sup_{\|\nu\| \leqslant B \left(\frac{m}{n} \right)^{\frac{1}{2}}} \left| \mathbb{P}_n \boldsymbol{\lambda}^\mathsf{T} \left(\eta_t(\nu) - \mathbb{E} \left[\eta_t(\nu) \middle| \mathcal{F}_{t-1} \right] \right) \right| = o_p \left(\frac{1}{\sqrt{mn}} \right).
$$

The considered process is a martingale difference sequence. The next step is to divide the ball defined as $\left\{\boldsymbol{\nu} \in \mathbb{R}^{2(m+1)+K+2} : \|\boldsymbol{\nu}\| \leqslant B(\frac{m}{n})^{\frac{1}{2}}\right\}$ in equation (3.18) into cubes $\mathcal{C}_j \subset \mathbb{R}^{2(m+1)+K+2}$ centered at ν_j and with side-length $m^{\frac{1}{2}} n^{-\frac{5}{2}}$. The resulting cardinality for $2(m+1)+K+2$ dimensions is then $N(n) := \|\{\mathcal{C}_j\}\| = (2n)^{2(m+1)+K+2}$. Now then for each $k \in \mathcal{I}_{1,K}$ the term $\eta_t(\nu)$ can be bounded by

$$\eta_t(\nu) \leqslant \eta_t(\nu_j) + b_{k,t}(\nu_j)\mathbf{x}_{t,k}(\boldsymbol{a}_0)$$

and similarly

$$\eta_t(\nu) \geqslant \eta_t(\nu_j) + \left(b_{k,t}(\nu_j) - d_{k,t}(\nu_j)\right)\mathbf{x}_{t,k}(\boldsymbol{a}_0),$$

with $b_{k,t}(\nu_j)\mathbf{x}_{t,k}(\boldsymbol{a}_0)$ and $d_{k,t}(\nu_j)\mathbf{x}_{t,k}(\boldsymbol{a}_0)$ being the process η_t evaluated at the maximum possible distance on each axis from the center to the boundary of the cube and the maximum possible distance on each axis between the boundaries of the cube $\mathcal{C}_j$, respectively:

$$\begin{aligned} b_{k,t}(\nu_j) &= \mathbb{1}\left\{u_t < (\boldsymbol{a}_0 + \nu_j)^{\mathsf{T}} \mathbf{x}_{t,k}(\boldsymbol{a}_0)\right\} \\ &\quad - \mathbb{1}\left\{u_t < (\boldsymbol{a}_0 + \nu_j)^{\mathsf{T}} \mathbf{x}_{t,k}(\boldsymbol{a}_0) + B\left(n^{\frac{1}{2}} m^{-\frac{5}{2}}\right) \|\mathbf{x}_{t,k}(\boldsymbol{a}_0)\|\right\}, \end{aligned}$$

$$\begin{aligned} d_{\tau,t}(\nu_j) &= \mathbb{1}\left\{u_t < (\boldsymbol{a}_0 + \nu_j)^{\mathsf{T}} \mathbf{x}_{t,k}(\boldsymbol{a}_0) + B\left(n^{\frac{1}{2}} m^{-\frac{5}{2}}\right) \|\mathbf{x}_{t,k}(\boldsymbol{a}_0)\|\right\} \\ &\quad - \mathbb{1}\left\{u_t < (\boldsymbol{a}_0 + \nu_j)^{\mathsf{T}} \mathbf{x}_{t,k}(\boldsymbol{a}_0) - B\left(n^{\frac{1}{2}} m^{-\frac{5}{2}}\right) \|\mathbf{x}_{t,k}(\boldsymbol{a}_0)\|\right\}. \end{aligned}$$

Taking expectations of the latter and subtracting it from the first one, implies that for all $\nu \in \mathcal{C}_j$, for all t, and for all k it holds that

$$\begin{aligned} (\eta_t(\nu) - \mathbb{E}[\eta_t(\nu)|\mathcal{F}_{t-1}]) &\leqslant (\eta(\nu_j) - \mathbb{E}[\eta_t(\nu_j)|\mathcal{F}_{t-1}]) \\ &\quad + (b_{k,t}(\nu_j) - \mathbb{E}[b_{k,t}(\nu_j)|\mathcal{F}_{t-1}])\,\mathbf{x}_{t,k}(\boldsymbol{a}_0) + \mathbb{E}[d_{k,t}(\nu_t)|\mathcal{F}_{t-1}]\,\mathbf{x}_{t,k}(\boldsymbol{a}_0), \end{aligned}$$

which implies that

$$\sup_{\|\nu\| \leqslant B\left(\frac{m}{n}\right)^{\frac{1}{2}}} \left|\mathbb{P}_n \boldsymbol{\lambda}^{\mathsf{T}} \left(\eta_t(\nu) - \mathbb{E}[\eta_t(\nu)|\mathcal{F}_{t-1}]\right)\right|$$

$$\leqslant \max_{j \in \mathcal{I}_{1,N(n)}} \left|\mathbb{P}_n \left\|\boldsymbol{\lambda}^{\mathsf{T}} \mathbf{x}_{t,k}(\boldsymbol{a}_0)\right\| (b_{k,t}(\nu_j) - \mathbb{E}[b_{k,t}(\nu_j)|\mathcal{F}_{t-1}])\right| \tag{3.23}$$

$$+ \max_{j \in \mathcal{I}_{1,N(n)}} \left|\mathbb{P}_n \left\|\boldsymbol{\lambda}^{\mathsf{T}} \mathbf{x}_{t,k}(\boldsymbol{a}_0)\right\| \mathbb{E}[d_{k,t}(\nu_t)|\mathcal{F}_{t-1}]\right| \tag{3.24}$$

$$+ \max_{j \in \mathcal{I}_{1,N(n)}} \left|\mathbb{P}_n \boldsymbol{\lambda}^{\mathsf{T}} \left(\eta(\nu_j) - \mathbb{E}[\eta_t(\nu_j)|\mathcal{F}_{t-1}]\right)\right|. \tag{3.25}$$

Expressions (3.23) and (3.24) are similar to Theorem 3.1 in Welsh [1989] who show that they are of negligible stochastic order for a general class of M-estimators and linear specifications, whereas $\mathbf{x}_{t,k}(\boldsymbol{a})$ here is non-linear, however it is evaluated at the true parameter $\boldsymbol{a}_0$. Since our problem is piece-wise linear with a bounded transition function, second moments of the conditional volatility process and by Assumption 3.6 strictly positive and finite eigenvalues of $\mathbf{D}_{1,m,n}$, their results apply to our setting. Expression (3.25) is identical to Xiao & Koenker [2009], equation (A.5). The latter requires the exponential bound Assumption 3.8 to be of negligible stochastic order. For further details, we refer the reader to these citations.

Hence, with equation (3.21) being negligible, equation (3.19) can be written as

$$\mathbb{P}_n \boldsymbol{\lambda}^\mathsf{T} \varphi_t \left(\boldsymbol{a}_0 + B \left(\frac{m}{n}\right)^{\frac{1}{2}} \boldsymbol{\lambda} \right) = \mathbb{P}_n \boldsymbol{\lambda}^\mathsf{T} \varphi_t (\boldsymbol{a}_0) + B \left(\frac{m}{n}\right)^{\frac{1}{2}} \boldsymbol{\lambda}^\mathsf{T} \mathbf{D}_{1,n,m} \boldsymbol{\lambda} + o_p \left(\left(\frac{m}{n}\right)^{\frac{1}{2}} \right). \tag{3.26}$$

Whenever the right hand side of this equation exceeds zero, it is implied that so does the left hand side. This is true with probability tending to one since by Assumption 3.6 the following equation holds:

$$\inf_{\boldsymbol{\lambda} \in \mathbb{R}^{2(m+1)+K+2} : \|\boldsymbol{\lambda}\|=1} \left(\frac{m}{n}\right)^{-\frac{1}{2}} \mathbb{P}_n \boldsymbol{\lambda}^\mathsf{T} \varphi_t (\boldsymbol{a}_0) > -\frac{B}{2} \lambda_{n,\min} - o_p \left(\sqrt{mn}\right) < 0$$

noting that $\mathbb{P}_n \boldsymbol{\lambda}^\mathsf{T} \varphi_t (\boldsymbol{a}_0) = \mathcal{O}_p(\sqrt{m/n})$ and $\lambda_{n,\min} > 0$ as $n \to \infty$ and $B \to \infty$. □

Corollary 3.1 (First Stage Bahadur Representation). Let $s_k = f_\varepsilon(F_\varepsilon^{-1}(\tau_k))$, $\mathbf{s}^\mathsf{T} = [s_1, \dots, s_K]$, $\mathbf{q}^\mathsf{T} = [q_1, \dots, q_K]$. Also let $\boldsymbol{\alpha}^\Delta = \boldsymbol{\alpha}^\mathrm{I} - \boldsymbol{\alpha}^\mathrm{II}$ and $\bar{\boldsymbol{\alpha}}(\zeta) = G_t(\zeta)\, \boldsymbol{\alpha}^\mathrm{I} + (1 - G_t(\zeta))\boldsymbol{\alpha}^\mathrm{II}$. Then under Assumptions 3.1-3.8, it holds for $n \to \infty$

$$\sqrt{n} \begin{bmatrix} \widehat{\boldsymbol{\alpha}_n^\mathrm{I}} - \boldsymbol{\alpha}_0^\mathrm{I} \\ \widehat{\boldsymbol{\alpha}_n^\mathrm{II}} - \boldsymbol{\alpha}_0^\mathrm{II} \\ \widehat{\zeta_n} - \zeta_0 \end{bmatrix} \approx \left(\sum_{k=1}^K s_k q_k^2 \right)^{-1} \mathbf{D}_{n,m}^{-1} \frac{1}{\sqrt{n}} \sum_{t=m+1}^N \begin{bmatrix} G_t(\zeta)\, \boldsymbol{z}_t^m \\ 1 - G_t(\zeta)\, \boldsymbol{z}_t^m \\ \boldsymbol{z}_t^{m,\mathsf{T}} \boldsymbol{\alpha}_0^\Delta \frac{\partial G_t(\zeta)}{\partial \zeta} \end{bmatrix} \sum_{k=1}^K q_k \left(\mathbb{1}\left\{ u_t \leqslant F^{-1}_{u_t|\mathcal{F}_{t-1}}(\tau_k) \right\} - \tau_k \right)$$

where the approximation is up to a stochastically negligible sequence of order $(m/n)^{1/2}$ and where

$$\mathbf{D}_{n,m} := -\mathbb{E} \frac{1}{n} \sum_{t=m}^n \frac{1}{\sigma_t} \begin{bmatrix} \boldsymbol{\Omega}_{11} & \boldsymbol{\Omega}_{12} \\ \boldsymbol{\Omega}_{12}^\mathsf{T} & \boldsymbol{\Omega}_{22} \end{bmatrix}.$$

with respective blocks

$$\begin{aligned}
\boldsymbol{\Omega}_{11} &= \begin{bmatrix} G_t(\zeta_0)^2 & G_t(\zeta_0)\,(1 - G_t(\zeta_0)) \\ G_t(\zeta_0)\,(1 - G_t(\zeta_0)) & (1 - G_t(\zeta_0))^2 \end{bmatrix} \otimes \boldsymbol{z}_t^m \boldsymbol{z}_t^{m,\mathsf{T}}, \\
\boldsymbol{\Omega}_{12} &= \boldsymbol{z}_t^{m\mathsf{T}} \boldsymbol{\alpha}_0^\Delta \left(\begin{bmatrix} G_t(\zeta_0) \\ 1 - G_t(\zeta_0) \end{bmatrix} \otimes \boldsymbol{z}_t^m \frac{\partial G}{\partial \zeta_0^\mathsf{T}} \right), \\
\boldsymbol{\Omega}_{22} &= (\boldsymbol{z}_t^{m\mathsf{T}} \boldsymbol{\alpha}_0^\Delta)^2 \frac{\partial G}{\partial \zeta_0} \frac{\partial G}{\partial \zeta_0^\mathsf{T}}.
\end{aligned}$$

Proof. Let $\widehat{\nu} = \widehat{\boldsymbol{a}}_n - \boldsymbol{a}_0$ where $\widehat{\boldsymbol{a}}_n$ solves the objective function defined in equation (3.9) and substitute $B \left(\frac{m}{n}\right)^{\frac{1}{2}} \boldsymbol{\lambda}$ by this sequence in equation (3.26) to get

$$\mathbb{P}_n \boldsymbol{\lambda}^\mathsf{T} \varphi_t (\widehat{\boldsymbol{a}}_n) = \mathbb{P}_n \boldsymbol{\lambda}^\mathsf{T} \varphi_t (\boldsymbol{a}_0) + \boldsymbol{\lambda}^\mathsf{T} \mathbf{D}_{1,n,m} (\widehat{\boldsymbol{a}}_n - \boldsymbol{a}_0) + o_p \left(\left(\frac{m}{n}\right)^{\frac{1}{2}} \right).$$

By construction, the moment function on the left hand side is zero the estimate $\widehat{\boldsymbol{a}}_n$. Thus, we have for the right hand side for all $\lambda \in \mathbb{R}^m$ satisfying $\|\lambda\| = 1$:

$$\boldsymbol{\lambda}^\mathsf{T} \left[\mathbb{P}_n \varphi_t (\boldsymbol{a}_0) + \mathbf{D}_{1,n,m} (\widehat{\boldsymbol{a}}_n - \boldsymbol{a}_0) + o_p \left(\frac{1}{\sqrt{n}} \right) \right] = 0,$$

and thus the expression inside the bracket must be zero. We premultiply the latter by $\sqrt{n}$ and get the Bahadur representation for $\sqrt{n}(\widehat{\mathbf{a}}_n - \mathbf{a}_0)$, as well as the one for $\sqrt{n}(\widehat{\boldsymbol{\alpha}}_n - \boldsymbol{\alpha}_0)$ by only considering the first $2(m+1)$ and the last 2 rows.

Applying the central limit theorem [Ibragimov & Linnik, 1971, Theorem 18.5.3] to $\frac{1}{\sqrt{n}}\sum_{t=1}^{N}\varphi_t(\mathbf{a}_0)$, we have for any linear combination of the components of the $(2(m+1)+2)$-dimensional Bahadur representation:

$$\sqrt{n}\boldsymbol{\mu}^{\mathsf{T}}(\widehat{\boldsymbol{\alpha}}_n - \boldsymbol{\alpha}_0) \rightsquigarrow \mathcal{N}\left(0\,,\ \frac{\sum\limits_{k=1}^{K}\sum\limits_{k'=1}^{K} q_k q_{k'}(\tau_k \wedge \tau_{k'})(1-\tau_k \vee \tau_{k'})}{\left(\sum\limits_{k=1}^{K} q_k^2 f_\varepsilon\left(F_\varepsilon^{-1}(\tau_k)\right)\right)^2}\boldsymbol{\mu}^{\mathsf{T}}\mathbf{D}_{n,m}^{-1}\boldsymbol{\mu}\right) \tag{3.27}$$

where $\boldsymbol{\mu} \in \mathbb{R}^{2(m+1)+2}$. For this we require the moment and mixing conditions stated in Assumption 3.5 ensuring that we have $(2+\delta)$ moments of our data and mixing coefficient satisfying $\beta_s \to 0$ and $\sum\limits_{s=1}^{\infty}\beta_s^{\delta/(2+\delta)} < +\infty$. □

For the second stage, the directional derivative of the objective function as in equation (3.11) is re-defined as

$$g_n(\theta, \boldsymbol{\alpha}) = g_n\left((\theta^{\mathrm{I}}, \theta^{\mathrm{II}}, \zeta), (\boldsymbol{\alpha}^{\mathrm{I}}, \boldsymbol{\alpha}^{\mathrm{II}}, \zeta)\right) := \mathbb{P}_n \begin{bmatrix} \boldsymbol{z}_t(\boldsymbol{\alpha})G_t(\zeta) \\ \boldsymbol{z}_t(\boldsymbol{\alpha})(1-G_t(\zeta)) \\ \theta^{\Delta\mathsf{T}}\boldsymbol{z}_t(\boldsymbol{\alpha})\frac{\partial G_t(\zeta)}{\partial \zeta} \end{bmatrix} \left(\tau - \mathbb{1}\left\{u_t \leqslant [\theta^{\mathrm{I}}, \theta^{\mathrm{II}}]^{\mathsf{T}}\boldsymbol{z}_t(\boldsymbol{\alpha}, \zeta)\right\}\right)$$

and the one for the population, using the law of iterative expectations, as

$$g(\theta, \boldsymbol{\alpha}) = g\left((\theta^{\mathrm{I}}, \theta^{\mathrm{II}}, \zeta), (\boldsymbol{\alpha}^{\mathrm{I}}, \boldsymbol{\alpha}^{\mathrm{II}}, \zeta)\right) := \mathbb{E}\left[\begin{bmatrix} \boldsymbol{z}_t(\boldsymbol{\alpha})G_t(\zeta) \\ \boldsymbol{z}_t(\boldsymbol{\alpha})(1-G_t(\zeta)) \\ \theta^{\Delta\mathsf{T}}\boldsymbol{z}_t(\boldsymbol{\alpha})\frac{\partial G_t(\zeta)}{\partial \zeta} \end{bmatrix} \left(\tau - F_{u|\mathcal{F}_{t-1}}\left([\theta^{\mathrm{I}}, \theta^{\mathrm{II}}]^{\mathsf{T}}\boldsymbol{z}_t(\boldsymbol{\alpha}, \zeta)\right)\right)\right].$$

Note that we re-estimate ζ and as such consider only the moments for the ζ parameter as a part of θ. Whenever we consider the parameter ζ related to the first stage (as part of $\boldsymbol{\alpha}$), it will be explicitly mentioned. Due to Assumption 3.3, the latter is differentiable such that the partial derivatives with respect to the two parameter vectors θ and $\boldsymbol{\alpha}$ evaluated at the true parameters are given by $\Gamma_\theta(\theta, \boldsymbol{\alpha}; \mu) :=$

$$-\frac{f_\varepsilon\left(F_\varepsilon^{-1}(\tau)\right)}{\sigma_\varepsilon}\int \left[\begin{array}{c:c} \begin{bmatrix} G_t(\zeta)^2 & G_t(\zeta)(1-G_t(\zeta)) \\ G_t(\zeta)(1-G_t(\zeta)) & (1-G_t(\zeta))^2 \end{bmatrix} \otimes \boldsymbol{z}_t(\boldsymbol{\alpha})\boldsymbol{z}_t(\boldsymbol{\alpha})^{\mathsf{T}} & \theta^{\Delta\mathsf{T}}\boldsymbol{z}_t(\boldsymbol{\alpha})\begin{bmatrix} G_t(\zeta) \\ 1-G_t(\zeta) \end{bmatrix} \otimes \boldsymbol{z}_t(\boldsymbol{\alpha})\frac{\partial G_t(\zeta)}{\partial \zeta^{\mathsf{T}}} \\ \hdashline & \left(\theta^{\Delta\mathsf{T}}\boldsymbol{z}_t(\boldsymbol{\alpha})\right)^2 \frac{\partial G_t(\zeta)}{\partial \zeta}\frac{\partial G_t(\zeta)}{\partial \zeta^{\mathsf{T}}} \end{array}\right] d\mu \tag{3.28}$$

and derivatives with respect to $\boldsymbol{\alpha}$

$$\Gamma_\alpha(\theta, \boldsymbol{\alpha}; \mu) := -\frac{f_\varepsilon\left(F_\varepsilon^{-1}(\tau)\right)}{\sigma_\varepsilon}\int \begin{bmatrix} \boldsymbol{z}_t(\boldsymbol{\alpha})G_t(\zeta) \\ \boldsymbol{z}_t(\boldsymbol{\alpha})(1-G_t(\zeta)) \\ \theta^{\Delta\mathsf{T}}\boldsymbol{z}_t(\boldsymbol{\alpha})\frac{\partial G_t(\zeta)}{\partial \zeta} \end{bmatrix} \theta^{\mathsf{T}} \begin{bmatrix} \begin{bmatrix} G_t(\zeta) \\ 1-G_t(\zeta) \end{bmatrix} \otimes \begin{bmatrix} 0_{p+1,m} \\ [L^1, \ldots, L^q]^{\mathsf{T}} \otimes \boldsymbol{z}_t^m \end{bmatrix} \\ \boldsymbol{\iota}_q^{\mathsf{T}}[L^1, \ldots, L^q]^{\mathsf{T}} \otimes \boldsymbol{z}_t^{m\mathsf{T}}\boldsymbol{\alpha}^{\Delta}\frac{\partial G}{\partial \zeta} \end{bmatrix} d\mu,$$

where L is the lag operator. The expectations $\Gamma_{\theta,0} = \Gamma_\theta(\theta_0, \boldsymbol{\alpha}_0; \mathbf{P})$, $\Gamma_{\alpha,0} := \Gamma_\alpha(\theta_0, \boldsymbol{\alpha}_0; \mathbf{P})$ well-defined and exist by Assumptions 3.1-3.5. In addition to this we have that $\Gamma_{\theta,0}$ is positive definite and has full rank which follows from the fact that our GARCH process is invertible by Assumption 3.2 and Assumption 3.6 which implies that $\Gamma_\theta = \sum_{m=1}^{\infty} \Psi(\boldsymbol{\alpha})^\top \mathbf{D}_{n,m}\Psi(\boldsymbol{\alpha})$ has full rank as well, where Ψ is a matrix that constructs z_t^{pq} from z_t^m. In addition to this let $\Gamma_{\theta,n}(\theta,\boldsymbol{\alpha}) = \Gamma_\theta(\theta,\boldsymbol{\alpha};\mathbb{P}_n)$ and $\Gamma_{\alpha,n}(\theta,\boldsymbol{\alpha}) := \Gamma_\alpha(\theta,\boldsymbol{\alpha};\mathbb{P}_n)$ be their sample analogues.

Theorem 3.2 (Second Stage Consistency)**.** Under Assumptions 3.1-3.8, the second-stage estimator is $\sqrt{n}$-consistent, that is, for $n \to \infty$ and $\tau \in (0,1)$

$$\left\|\widehat{\theta}_n(\tau) - \theta_0(\tau)\right\| = \mathcal{O}_p(n^{-\frac{1}{2}}).$$

Proof. It needs to be shown that $\left\|\widehat{\theta}_n(\tau) - \theta_0(\tau)\right\| = \mathcal{O}_p\left(n^{-\frac{1}{2}}\right)$. For this, note that $g(\theta,\boldsymbol{\alpha})$ is differentiable for any θ. Thus, the first order Taylor series expansion of around $\widehat{\theta}_n(\tau)$ can be applied

$$g\left(\widehat{\theta}_n(\tau),\boldsymbol{\alpha}_0\right) - g\left(\theta_0(\tau),\boldsymbol{\alpha}_0\right) = \Gamma_{\theta,0}\left(\widehat{\theta}_n(\tau) - \theta_0(\tau)\right).$$

Taking norms, a bound for the right hand side is obtained:

$$\left\|g\left(\widehat{\theta}_n(\tau),\boldsymbol{\alpha}_0\right) - g\left(\theta_0(\tau),\boldsymbol{\alpha}_0\right)\right\| \geqslant \lambda_{\min}\left(\Gamma_{\theta,0}\right)\left\|\left(\widehat{\theta}_n(\tau) - \theta_0(\tau)\right)\right\|,$$

with $\lambda_{\min}\left(\Gamma_{\theta,0}\right)$ being the the smallest eigenvalue of $\Gamma_{\theta,0}$ which is strictly positive as argued above. Since $g\left(\theta_0(\tau),\boldsymbol{\alpha}_0\right) = 0$, it is sufficient to show that $\left\|g\left(\widehat{\theta}_n(\tau),\boldsymbol{\alpha}_0\right)\right\| = \mathcal{O}_p\left(n^{-\frac{1}{2}}\right)$. Using the triangle inequality it follows that

$$\begin{aligned}
\left\|g\left(\widehat{\theta}_n(\tau),\boldsymbol{\alpha}_0)\right)\right\| &\leqslant \left\|g\left(\widehat{\theta}_n(\tau),\boldsymbol{\alpha}_0\right) - g\left(\widehat{\theta}_n(\tau),\widehat{\boldsymbol{\alpha}}_n\right)\right\| + \left\|g\left(\widehat{\theta}_n(\tau),\widehat{\boldsymbol{\alpha}}_n\right)\right\| && (3.29)\\
&\leqslant \left\|g\left(\widehat{\theta}_n(\tau),\boldsymbol{\alpha}_0\right) - g\left(\widehat{\theta}_n(\tau),\widehat{\boldsymbol{\alpha}}_n\right)\right\| && (3.30)\\
&\quad + \left\|g\left(\widehat{\theta}_n(\tau),\widehat{\boldsymbol{\alpha}}_n\right) - g\left(\theta_0(\tau),\boldsymbol{\alpha}_0\right) - g_n\left(\widehat{\theta}_n(\tau),\widehat{\boldsymbol{\alpha}}_n\right) + g_n\left(\theta_0(\tau),\boldsymbol{\alpha}_0\right)\right\| && (3.31)\\
&\quad + \left\|g_n\left(\widehat{\theta}_n(\tau),\widehat{\boldsymbol{\alpha}}_n\right)\right\| && (3.32)\\
&\quad + \left\|g_n\left(\theta_0(\tau),\boldsymbol{\alpha}_0\right)\right\|, && (3.33)
\end{aligned}$$

where $g\left(\theta_0(\tau),\boldsymbol{\alpha}_0\right) = 0$ was subtracted within the second norm (3.31). By the central limit theorem [Ibragimov & Linnik, 1971, Theorem 18.5.3], the existence of the $(2+\delta)$ moment of z_t^m, $z_t^m G_t(\zeta_0)$ and $z_t^m \partial G_t(\zeta)\zeta$, respectively, the boundedness of both $G_t(\zeta_0)$ and $\frac{\partial G_t(\zeta_0)}{\partial \zeta}$ the expression (3.33) is tight and we have $\left\|g_n\left(\theta_0(\tau),\boldsymbol{\alpha}_0\right)\right\| = \mathcal{O}_p\left(n^{-\frac{1}{2}}\right)$. The remaining equations (3.30), (3.31) and (3.32) can again be analyzed separately. Starting with the first term, again using the triangle inequality, and changing the signs within the norm, equation (3.30) can be bounded by

$$\begin{aligned}
\left\|g\left(\widehat{\theta}_n(\tau),\boldsymbol{\alpha}_0\right) - g\left(\widehat{\theta}_n(\tau),\widehat{\boldsymbol{\alpha}}_n\right)\right\| &\leqslant \left\|g\left(\widehat{\theta}_n(\tau),\widehat{\boldsymbol{\alpha}}_n\right) - g\left(\widehat{\theta}_n(\tau),\boldsymbol{\alpha}_0\right) - \Gamma_{\alpha,n}\left(\widehat{\theta}_n(\tau),\boldsymbol{\alpha}_0\right)\left(\widehat{\boldsymbol{\alpha}}_n - \boldsymbol{\alpha}_0\right)\right\| && (3.34)\\
&\quad + \left\|\Gamma_{\alpha,n}\left(\widehat{\theta}_n(\tau),\boldsymbol{\alpha}_0\right)\left(\widehat{\boldsymbol{\alpha}}_n - \boldsymbol{\alpha}_0\right) - \Gamma_{\alpha,n}\left(\theta_0(\tau),\boldsymbol{\alpha}_0\right)\left(\widehat{\boldsymbol{\alpha}}_n - \boldsymbol{\alpha}_0\right)\right\| && (3.35)\\
&\quad + \left\|\Gamma_{\alpha,n}\left(\theta_0(\tau),\boldsymbol{\alpha}_0\right)\left(\widehat{\boldsymbol{\alpha}}_n - \boldsymbol{\alpha}_0\right)\right\|
\end{aligned}$$

Applying the Taylor series expansion of $g\left(\widehat{\theta}_n(\tau),\widehat{\boldsymbol{\alpha}}_n\right)$ around $\boldsymbol{\alpha}_0$ in equation (3.34) and using the fact that $\Gamma_{\alpha,n}$ is Lipschitz in α due to the fact that $\frac{\partial G_t(\zeta_0)}{\partial \zeta}$ is Lipschitz and σ_t is bounded due to the fact that it is invertible to an ARCH model by Assumption 3.2 and the boundedness of u_t by Assumption 3.4. Similarly for equation (3.35) we use that $\Gamma_{\alpha,n}\left(\widehat{\theta}_n(\tau), \alpha_0\right)$ is Lipschitz in $\theta_0(\tau)$, which has bounded parameter space. Thus, this reduces to

$$\left\|g\left(\widehat{\theta}_n(\tau), \alpha_0\right) - g\left(\widehat{\theta}_n(\tau), \widehat{\alpha}_n\right)\right\| \leqslant \mathcal{O}_p\left(\|\widehat{\alpha}_n - \alpha_0\|^2\right) + \mathcal{O}_p\left(\|\widehat{\alpha}_n - \alpha_0\| \left\|\widehat{\theta}_n(\tau) - \theta_0(\tau)\right\|\right) + \|\Gamma_{\alpha,0}\left(\widehat{\alpha}_n - \alpha_0\right)\| \tag{3.36}$$

$$= \|\Gamma_{\alpha,0}\left(\widehat{\alpha}_n - \alpha_0\right)\| \left(1 + o_p\left(1\right)\right), \tag{3.37}$$

where the the last term follows from the fact that $\boldsymbol{\Gamma}_{\alpha,n} \to \boldsymbol{\Gamma}_{\alpha,0}$ in probability by law of large numbers, for which we need the existence of the moments in Assumption 3.5, and Slutsky's lemma. In a next step, we analyze the remaining terms (3.31) and (3.32), for which we have to check the conditions for Lemma 4.2 in Chen [2008]. For this, let

$$m_\tau(z_t, \theta, \alpha) = \begin{bmatrix} z_t(\alpha) G_t(\zeta) \\ z_t(\alpha)(1 - G_t(\zeta)) \\ \theta^{\Delta T} z_t(\alpha) \frac{\partial G_t(\zeta)}{\partial \zeta} \end{bmatrix} \left(\tau - \mathbb{1}\left\{u_t \leqslant [\theta^{I}, \theta^{II}]^T z_t(\alpha, \zeta)\right\}\right) \tag{3.38}$$

such that $g_n(\theta, \alpha) = \mathbb{P}_n m_\tau(z_t(\alpha), \theta, \alpha)$ and $g(\theta, \alpha) = \mathbb{E} m_\tau(z_t, \theta, \alpha)$. Then if z_t is stationary which is true by Assumption 3.2 and 3.3, has β-mixing decay rate as in Assumption 3.5 (see e.g. Carrasco & Chen [2002], Meitz & Saikkonen [2008]), $\boldsymbol{\Theta}_2^\tau$ is a compact subset of $\mathbb{R}^{2(p+q+1)+2}$ and $\boldsymbol{\Theta}_1$ one of $\mathbb{R}^{2(m+1)} \times \mathbb{R} \times \mathbb{R}_+$ for each $m_{\tau, j \in \mathcal{I}_{2(p+q+1)+2}} := (m_\tau)_j$, we only have to verify that

$$\left(\mathbb{E}\left[\sup_{((\theta'', \zeta''), (\alpha', \zeta')) \in \mathfrak{U}_\delta(\theta_0(\tau), \alpha_0)} \left|m_{\tau,j}(z_t, \theta'', \zeta'', \alpha', \zeta') - m_{\tau,j}(z_t, \theta_0, \zeta_0, \alpha_0, \zeta_0)\right|^r\right]\right)^{\frac{1}{r}} \leqslant K_j \delta^{s_j}$$

for some s_j which is bound by the degree of smoothness of $G_t(\zeta)$, some constant $K_j > 0$ and for $r = 2 + \delta$, satisfying the restriction in Assumption 3.5 to claim that:

$$\sup_{(\theta', \alpha') \in \mathfrak{U}_\delta(\theta_0(\tau), \alpha_0)} \left\|g\left(\theta', \alpha'\right) - g\left(\theta_0(\tau), \alpha_0\right) - g_n\left(\theta', \alpha'\right) + g_n\left(\theta_0(\tau), \alpha_0\right)\right\| = o_p\left(n^{-\frac{1}{2}}\right).$$

with $\mathfrak{U}_\delta := \{(\theta, \alpha) \in \boldsymbol{\Theta}_2^\tau \times \boldsymbol{\Theta}_1 : \|\theta - \theta_0(\tau)\| \leqslant \delta, \|\alpha - \alpha_0\| \leqslant \delta\}$ by Lemma 4.2 in Chen [2008]. As discussed in Chen [2008] we need $m_{\tau,j}$ to be a member of the function class with bracketing number that satisfies $\int_0^\infty \sqrt{\log N(\epsilon^{1/s_j}, \mathcal{H}, \|\cdot\|_{\mathcal{H}})} d\epsilon < \infty$ where the degree of smoothness sastisfies $d \geqslant 2/(2s_j)$ with $s_j = 1$. Alternatively we can make use of the class of monotone functions which is sufficient for the former condition. A detailed discussion of this can be found in Chen [2008]. This implies that we either need continuity of $\frac{\partial G_t(\zeta)}{\partial \zeta}$ [van der Vaart & Wellner, 1996, Theorem 2.7.1] or monotonicity [van der Vaart & Wellner, 1996, Theorem 2.7.5] of $\frac{\partial G_t(\zeta)}{\partial \zeta}$ with respect to ξ_t. Assumption 3.1 ensures that the transition function belongs to one of these classes. Note that, the continuity condition holds for the derivative of the logistic transition function but not for the linear one, whereas monotonicity only holds for the derivative of the linear one.

The uniform boundedness relative to the L^r-norm of the distance between (3.38) evaluated at any two parameter values within a neighborhood of the true parameters can be shown as follows.

By definition we have

$$z_{t,j}(\theta'', \zeta'', \alpha', \zeta') = \begin{cases} G(\zeta'')z_{t,j}(\alpha', \zeta') & \text{if} \quad 0 < j \leqslant p+q+1 \\ (1-G(\zeta''))z_{t,j-(p+q+1)}(\alpha', \zeta') & \text{if} \quad p+q+1 < j \leqslant 2(p+q+1) \\ \theta''^{\Delta T} z_t(\alpha', \zeta') \frac{\partial G(\zeta'', \eta'')}{\partial \zeta} & \text{if} \quad j = 2(p+q+1)+1 \\ \theta''^{\Delta T} z_t(\alpha', \zeta') \frac{\partial G(\zeta'', \eta'')}{\partial \eta} & \text{if} \quad j = 2(p+q+1)+2 \end{cases}.$$

In addition to this we have,

$$\begin{aligned} & \left| m_{\tau,j}(z_t, \theta'', \zeta'', \alpha', \zeta') - m_{\tau,j}(z_t, \theta_0, \zeta_0, \alpha_0, \zeta_0) \right|^r \\ & \leqslant \tau \left| z_{t,j}(\theta'', \zeta'', \alpha', \zeta') - z_{t,j}(\theta'', \zeta'', \alpha', \zeta') \right|^r && (3.39) \\ & + \left| z_{t,j}(\theta'', \zeta'', \alpha', \zeta') \mathbb{1}\left\{ u_t \leqslant \theta''^T z_t(\alpha', \zeta', \zeta'') \right\} - z_{t,j}(\theta_0, \zeta_0, \alpha_0, \zeta_0) \mathbb{1}\left\{ u_t \leqslant \theta_0^T z_t(\alpha_0, \zeta_0, \zeta_0) \right\} \right|^r. && (3.40) \end{aligned}$$

We start by expanding equation (3.39) and bounding each term individually:

$$\begin{aligned} \tau \mathbb{E}\left| z_{t,j}(\theta'', \zeta'', \alpha', \zeta') - z_{t,j}(\theta_0, \zeta_0, \alpha_0, \zeta_0) \right|^r \leqslant\; & \tau \mathbb{E}\left| z_{t,j}(\theta'', \zeta'', \alpha', \zeta') - z_{t,j}(\theta_0, \zeta'', \alpha', \zeta') \right|^r \\ & + \tau \mathbb{E}\left| z_{t,j}(\theta_0, \zeta'', \alpha', \zeta') - z_{t,j}(\theta_0, \zeta'', \alpha_0, \zeta') \right|^r \\ & + \tau \mathbb{E}\left| z_{t,j}(\theta_0, \zeta'', \alpha_0, \zeta') - z_{t,j}(\theta_0, \zeta_0, \alpha_0, \zeta') \right|^r \\ & + \tau \mathbb{E}\left| z_{t,j}(\theta_0, \zeta_0, \alpha_0, \zeta') - z_{t,j}(\theta_0, \zeta_0, \alpha_0, \zeta_0) \right|^r. \end{aligned}$$

Given that $\left\|\theta'' - \theta\right\| \leqslant \delta$, a bound for the first term can be obtained by noting that we have finite $(2+\delta)$ moments of u_t (for all t) and finite derivatives of the transition function G by Assumptions 3.1 and 3.6. such that $\tau^r \mathbb{E}\left|(\theta''^{\Delta} - \theta_0^{\Delta})^T z_{t,j}(\alpha', \zeta') \frac{\partial G(\zeta'', \eta'')}{\partial \bullet}\right|^r \leqslant K_{1,1,j}\delta^r$. For $\|\alpha' - \alpha\| \leqslant \delta$, a bound – denoted by $K_{1,2,j}\delta^r$ – follows immediately from the linearity of $z_{t,j}(\theta'', \zeta'', \alpha', \zeta')$ with respect to α and finite second moments of u_t. For the transition parameters $\left\|\zeta'' - \zeta\right\| \leqslant \delta$ and $\left\|\zeta' - \zeta\right\| \leqslant \delta$, we additionally need differentiability of G and the Lipschitz continuity and boundedness of $\frac{\partial G}{\partial \zeta}$, to get upper bounds denoted by $K_{1,3,j}$ and $K_{1,4,j}$, respectively.

Putting all these terms together we get the bound with constant $K_{1,j} = 4 \sup_{\iota} K_{1,\iota,j}$

$$\tau^r \mathbb{E}\left| z_{t,j}(\theta'', \zeta''\alpha', \zeta') - z_{t,j}(\theta_0, \alpha_0, \zeta_0, \zeta_0) \right|^r \leqslant K_{1,j}\delta^r. \tag{3.41}$$

For the second term (3.40) we note that

$$\begin{aligned} & \left| z_{t,j}(\theta'', \zeta'', \alpha', \zeta') \mathbb{1}\left\{ u_t \leqslant \theta''^T z_t(\alpha', \zeta', \zeta'') \right\} - z_{t,j}(\theta_0, \zeta_0, \alpha_0, \zeta_0) \mathbb{1}\left\{ u_t \leqslant \theta_0^T z_t(\alpha_0, \zeta_0, \zeta_0) \right\} \right|^r \\ & \leqslant \left| z_{t,j}(\theta'', \zeta'', \alpha', \zeta') \left(\mathbb{1}\left\{ u_t \leqslant \theta''^T z_t(\alpha', \zeta', \zeta'') \right\} - \mathbb{1}\left\{ u_t \leqslant \theta_0^T z_t(\alpha_0, \zeta_0, \zeta_0) \right\} \right) \right|^r && (3.42) \\ & + \left| \left(z_{t,j}(\theta'', \zeta'', \alpha', \zeta') - z_{t,j}(\theta_0, \zeta_0, \alpha_0, \zeta_0) \right) \mathbb{1}\left\{ u_t \leqslant \theta_0^T z_t(\alpha_0, \zeta_0, \zeta_0) \right\} \right|^r. && (3.43) \end{aligned}$$

Whereas the bound of the expectation of term (3.43) follows from (3.41), we need to take care of equation

(3.42). Let $z'_{t,j} = z_{t,j}(\theta'', \zeta'', \alpha', \zeta')$. Then we have

$$
\begin{aligned}
&\sup_{((\theta'',\zeta''),(\alpha',\zeta'))\in\mathfrak{U}_\delta(\theta_0(\tau),\alpha_0)} \mathbb{E}\left|z'_{t,j}\left(\mathbb{1}\left\{u_t \leqslant \theta''^{\mathsf{T}} z_t(\alpha',\zeta',\zeta'')\right\} - \mathbb{1}\left\{u_t \leqslant \theta_0^{\mathsf{T}} z_t(\alpha_0,\zeta_0,\zeta_0)\right\}\right)\right|^r \qquad (3.44)\\
\leqslant &\sup_{((\theta'',\zeta''),(\alpha',\zeta'))\in\mathfrak{U}_\delta(\theta_0(\tau),\alpha_0)} \Big\{\mathbb{E}\left|z'_{t,j}\right|^r\left(\mathbb{1}\left\{u_t \leqslant \theta''^{\mathsf{T}} z_t(\alpha',\zeta',\zeta'')\right\} - \mathbb{1}\left\{u_t \leqslant \theta_0^{\mathsf{T}} z_t(\alpha',\zeta',\zeta'')\right\}\right)\\
&\qquad + \mathbb{E}\left|z'_{t,j}\right|^r\left(\mathbb{1}\left\{u_t \leqslant \theta_0^{\mathsf{T}} z_t(\alpha',\zeta',\zeta'')\right\} - \mathbb{1}\left\{u_t \leqslant \theta_0^{\mathsf{T}} z_t(\alpha_0,\zeta',\zeta'')\right\}\right)\\
&\qquad + 2\mathbb{E}\left|z'_{t,j}\right|^r\left(\mathbb{1}\left\{u_t \leqslant \theta_0^{\mathsf{T}} z_t(\alpha_0,\zeta',\zeta'')\right\} - \mathbb{1}\left\{u_t \leqslant \theta_0^{\mathsf{T}} z_t(\alpha_0,\zeta',\zeta_0)\right\}\right)\\
&\qquad + 2\mathbb{E}\left|z'_{t,j}\right|^r\left(\mathbb{1}\left\{u_t \leqslant \theta_0^{\mathsf{T}} z_t(\alpha_0,\zeta',\zeta_0)\right\} - \mathbb{1}\left\{u_t \leqslant \theta_0^{\mathsf{T}} z_t(\alpha_0,\zeta_0,\zeta_0)\right\}\right)\Big\}
\end{aligned}
$$

where the inequality follows from the triangle inequality and the fact that $\omega \mapsto \mathbb{1}\{u_t \leqslant \omega\}$ is monotone in ω and which is in turn linear in θ and α, an in addition to this $\mathbb{1} \circ G$ monotone in the first parameter (location) of ζ, namely ζ and piece-wise monotone – increasing over half of the domain and decreasing over the other half – in its second parameter η (scale).

We then obtain another upper bound for the right hand side of equation (3.44) by applying the law of iterated expectations. By extension this is also an upper bound for the left hand side of equation (3.44) and we have:

$$
\begin{aligned}
&\sup_{((\theta'',\zeta''),(\alpha',\zeta'))\in\mathfrak{U}_\delta(\theta_0(\tau),\alpha_0)} \mathbb{E}\left|z'_{t,j}\left(\mathbb{1}\left\{u_t \leqslant \theta''^{\mathsf{T}} z_t(\alpha',\zeta',\zeta'')\right\} - \mathbb{1}\left\{u_t \leqslant \theta_0^{\mathsf{T}} z_t(\alpha_0,\zeta_0,\zeta_0)\right\}\right)\right|^r\\
\leqslant &\sup_{((\theta'',\zeta''),(\alpha',\zeta'))\in\mathfrak{U}_\delta(\theta_0(\tau),\alpha_0)} \mathbb{E}\Big\{\left|z'_{t,j}\right|^r f_{u_t|\mathcal{F}_{t-1}}\left(\tilde{\phi}_1\right) z_t(\alpha',\zeta',\zeta'')^{\mathsf{T}}\left(\theta''-\theta_0\right)\\
&+ \left|z'_{t,j}\right|^r f_{u_t|\mathcal{F}_{t-1}}\left(\tilde{\phi}_2\right)\left(G(\zeta'')\sum_{j=1}^{q}\theta^{I}_{p+j+1}L^j z_t(\zeta') + (1-G(\zeta''))\sum_{j=1}^{q}\theta^{II}_{p+j+1}L^j z_t(\zeta')\right)^{\mathsf{T}}(\alpha'-\alpha_0)\\
&+ 2\left|z'_{t,j}\right|^r f_{u_t|\mathcal{F}_{t-1}}\left(\tilde{\phi}_3\right)\left(z_t(\alpha_0,\zeta',\zeta'')^{\mathsf{T}}\theta_0^{\Delta}\frac{\partial G(\zeta_0)}{\partial\zeta}\right)^{\mathsf{T}}(\zeta''-\zeta_0)\\
&+ 2\left|z'_{t,j}\right|^r f_{u_t|\mathcal{F}_{t-1}}\left(\tilde{\phi}_4\right)\\
&\times\left(G(\zeta'')\sum_{j=1}^{q}\theta^{I}_{p+j+1}L^j z_t(\zeta')^{\mathsf{T}}\alpha^{\Delta}\frac{\partial L^j G_t(\zeta_0)}{\partial\zeta} + (1-G(\zeta''))\sum_{j=1}^{q}\theta^{II}_{p+j+1}L^j z_t(\zeta')^{\mathsf{T}}\alpha^{\Delta}\frac{\partial L^j G_t(\zeta_0)}{\partial\zeta}\right)^{\mathsf{T}}(\zeta'-\zeta_0)\Big\}.
\end{aligned}
$$

The variables $\tilde{\phi}_j$ for $j \in \mathcal{I}_{1,4}$ refer to the elements of small neighborhoods of the respective parameters around which we applied the mean value theorem, for which we require the density $f_{u_t|\mathcal{F}_{t-1}}$ to exist and to be bounded which holds by Assumption 3.4. For the last term we require the bound on $\frac{\partial G_t(\zeta)}{\partial\zeta}$. We also use the fact that $z_{t,j}$ has finite $(2+\delta)$ moments by Assumption 3.5. Thus, for any $(\theta'', \zeta'', \alpha', \zeta)$ in a neighborhood of their true counterparts, there exists a $K_{2,j} > 0$ such that the original equation (3.40) can be bounded by

$$
\mathbb{E}\left|z_{t,j}(\theta'',\zeta'',\alpha',\zeta')\mathbb{1}\left\{u_t \leqslant \theta''^{\mathsf{T}} z_t(\alpha',\zeta',\zeta'')\right\} - z_{t,j}(\theta_0,\zeta_0,\alpha_0,\zeta_0)\mathbb{1}\left\{u_t \leqslant \theta_0^{\mathsf{T}} z_t(\alpha_0,\zeta_0,\zeta_0)\right\}\right|^r \leqslant K_{2,j}\delta^r
$$

and thus, by Lemma 4.2 in Chen [2008] the following expression holds:

$$\sup_{(\theta',\alpha')\in\mathfrak{U}_\delta(\theta_0(\tau),\alpha_0)} \frac{\sqrt{n}\left\|g(\theta,\alpha)-g(\theta_0(\tau),\alpha_0)-g_n(\theta,\alpha)+g_n(\theta_0(\tau),\alpha_0)\right\|}{1+\sqrt{n}\left(\|g(\theta,\alpha)\|+\|g_n(\theta,\alpha)\|\right)} = o_p\left(n^{-1/2}\right).$$

Thus,

$$\begin{aligned}
&\left\|g\left(\widehat{\theta}_n(\tau),\widehat{\alpha}_n\right)-g(\theta_0(\tau),\alpha_0)+g_n\left(\widehat{\theta}_n(\tau),\widehat{\alpha}_n\right)-g_n(\theta_0(\tau),\alpha_0)\right\| \\
&\leqslant o_p(1)\left(n^{-\frac{1}{2}}+\left\|g\left(\widehat{\theta}_n(\tau),\widehat{\alpha}_n\right)\right\|+\left\|g_n\left(\widehat{\theta}_n(\tau),\widehat{\alpha}_n\right)\right\|\right) \\
&\leqslant o(1)+o_p(1)\left(\left\|g\left(\widehat{\theta}_n(\tau),\alpha_0\right)\right\|(1+o_p(1))+\left\|g_n\left(\widehat{\theta}_n(\tau),\widehat{\alpha}_n\right)]\right\|\right),
\end{aligned} \tag{3.45}$$

where the last step follows from equations (3.36) and (3.37). Thus, by applying inequalities (3.36) and (3.45) to the original inequality (3.29) the following bound is obtained:

$$\begin{aligned}
\left\|g\left(\widehat{\theta}_n(\tau),\alpha_0\right)\right\| &\leqslant \mathcal{O}_p\left(\|\widehat{\alpha}_n-\alpha_0\|^2\right)+\mathcal{O}_p\left(\|\widehat{\alpha}_n-\alpha_0\|\left\|\widehat{\theta}_n(\tau)-\theta_0(\tau)\right\|\right)+\|\Gamma_{\alpha,0}(\widehat{\alpha}_n-\alpha_0)\| \\
&\quad+\left\|g_n\left(\widehat{\theta}_n(\tau),\widehat{\alpha}_n\right)\right\|+o_p(1)\left\|g\left(\widehat{\theta}_n(\tau),\alpha_0\right)\right\|+o_p(1)\left\|g_n\left(\widehat{\theta}_n(\tau),\widehat{\alpha}_n\right)\right\|.
\end{aligned}$$

This can be rearranged and reduced to

$$(1-o_p(1))\left\|g\left(\widehat{\theta}_n(\tau),\alpha_0\right)\right\| \leqslant (1+o_p(1))\left\|g_n\left(\widehat{\theta}_n(\tau),\widehat{\alpha}_n\right)\right\|+\mathcal{O}_p\left(n^{-\frac{1}{2}}\right)$$

where the first term on right hand side is the $\inf\limits_{\theta\in\Theta^{\tau}_{2}}\|g_n(\theta,\widehat{\alpha}_n)\|$, which we need to show is $\mathcal{O}_p\left(n^{-\frac{1}{2}}\right)$. Note that

$$\begin{aligned}
\|g_n(\theta,\widehat{\alpha}_n)\| &\leqslant \|g(\theta,\widehat{\alpha}_n)-g(\theta_0(\tau),\alpha_0)+g_n(\theta,\widehat{\alpha}_n)-g_n(\theta_0(\tau),\alpha_0)\| \\
&\quad+\|g(\theta,\widehat{\alpha}_n)-g(\theta,\alpha_0)\|+\|g(\theta,\alpha_0)\|+\|g_n(\theta_0(\tau),\alpha_0)\| \\
&\leqslant o_p(1)\left(\|g(\theta,\widehat{\alpha}_n)\|+\|g_n(\theta,\widehat{\alpha}_n)\|\right)+\|g(\theta,\alpha_0)\|+\mathcal{O}_p\left(n^{-\frac{1}{2}}\right),
\end{aligned}$$

where the last inequality follows from equation (3.45) and Corollary 3.1. Thus,

$$\begin{aligned}
(1-o_p(1))\|g_n(\theta,\widehat{\alpha}_n)\| &\leqslant o_p(1)\|g(\theta,\widehat{\alpha}_n)\|+\|g(\theta,\alpha_0)\|+\mathcal{O}_p\left(n^{-\frac{1}{2}}\right) \\
&\leqslant (1+o_p(1))\left(\|g(\theta,\alpha_0)\|+\|\Gamma_{\alpha,0}(\widehat{\alpha}_n-\alpha_0)\|\right)+\mathcal{O}_p\left(n^{-\frac{1}{2}}\right),
\end{aligned}$$

for which equation (3.37) is used. Consequently, this is implies for the respective infima

$$\|g_n(\theta,\widehat{\alpha}_n)\| \leqslant \|g(\theta,\alpha_0)\|+(1+o_p(1))\|\Gamma_{\alpha,0}(\widehat{\alpha}_n-\alpha_0)\|$$

and, using the fact that $\Gamma_{\alpha,0}(\widehat{\alpha}_n-\alpha_0)=\mathcal{O}_p(n^{-\frac{1}{2}})$ by equation (3.27), we have:

$$\lambda_{\min}(\Gamma_{\theta,0})\left\|\left(\widehat{\theta}_n(\tau)-\theta_0(\tau)\right)\right\| \leqslant \left\|g\left(\widehat{\theta}_n(\tau),\alpha_0\right)\right\| = \mathcal{O}_p\left(n^{-\frac{1}{2}}\right),$$

which completes the proof. □

Corollary 3.2. Under Assumptions 3.1-3.8, the following linearization holds:

$$g_n(\theta, \widehat{\alpha}_n) - g_n(\theta_0(\tau), \alpha_0) - \Gamma_{\theta,0}(\theta - \theta_0(\tau)) - \Gamma_{\alpha,0}(\widehat{\alpha}_n - \alpha_0) = o_p\left(n^{-\frac{1}{2}}\right). \tag{3.46}$$

Proof. By adding and subtracting, equation (3.46) can be rewritten as:

$$\begin{aligned}
&g_n(\theta, \widehat{\alpha}_n) - g_n(\theta_0(\tau), \alpha_0) - \Gamma_{\theta,0}(\theta - \theta_0(\tau)) - \Gamma_{\alpha,0}(\widehat{\alpha}_n - \alpha_0)\\
&\quad = g_n(\theta, \widehat{\alpha}_n) - g_n(\theta_0(\tau), \alpha_0) - \Gamma_{\theta,0}(\theta - \theta_0(\tau)) - \Gamma_{\alpha,0}(\widehat{\alpha}_n - \alpha_0)\\
&\quad + g(\theta, \widehat{\alpha}_n) - g(\theta, \widehat{\alpha}_n) + g(\theta_0(\tau), \alpha_0) - g(\theta_0(\tau), \alpha_0)\\
&\quad + g(\theta, \alpha_0) - g(\theta, \alpha_0) + \Gamma_{\alpha,n}(\theta, \alpha_0)(\widehat{\alpha}_n - \alpha_0) - \Gamma_{\alpha,n}(\theta, \alpha_0)(\widehat{\alpha}_n - \alpha_0)
\end{aligned}$$

Again taking norms, rearranging the terms on the right hand side, using the triangle inequality, the following bound is obtained for $\theta = \widehat{\theta}_n(\tau)$:

$$\begin{aligned}
&\left\|g_n\left(\widehat{\theta}_n, \widehat{\alpha}_n\right) - g_n(\theta_0(\tau), \alpha_0) - \Gamma_{\theta,0}(\widehat{\theta}_n - \theta_0(\tau)) - \Gamma_{\alpha,0}(\widehat{\alpha}_n - \alpha_0)\right\|\\
&\leqslant \left\|g_n\left(\widehat{\theta}_n(\tau), \widehat{\alpha}_n\right) - g_n(\theta_0(\tau), \alpha_0) - \left(g\left(\widehat{\theta}_n(\tau), \widehat{\alpha}_n\right) - g(\theta_0(\tau), \alpha_0)\right)\right\|\\
&+ \left\|g\left(\widehat{\theta}_n(\tau), \widehat{\alpha}_n\right) - g\left(\widehat{\theta}_n(\tau), \alpha_0\right)\right) - \Gamma_{\alpha,n}\left(\widehat{\theta}_n(\tau), \alpha_0\right)(\widehat{\alpha}_n - \alpha_0)\right\|\\
&+ \left\|g\left(\widehat{\theta}_n(\tau), \alpha_0\right) - g(\theta_0(\tau), \alpha_0) - \Gamma_{\theta,0}\left(\widehat{\theta}_n(\tau) - \theta_0(\tau)\right)\right\|\\
&+ \left\|\Gamma_{\alpha,n}\left(\widehat{\theta}_n(\tau), \alpha_0\right)(\widehat{\alpha}_n - \alpha_0) - \Gamma_{\alpha,0}(\widehat{\alpha}_n - \alpha_0)\right\| = o_p\left(n^{-\frac{1}{2}}\right),
\end{aligned}$$

where we use stochastic equicontinuity for the first term and reason along the lines of (3.34), (3.35), (3.36) using Lipschitz continuity of $\Gamma_{\alpha,0}$ and $\Gamma_{\theta,0}$, as well as the law of large numbers for $\Gamma_{\alpha,n}$ and $\sqrt{n}$-consistency of $\widehat{\theta}_n(\tau)$. □

Theorem 3.3 (Second Stage Asymptotic Normality)**.** If Assumptions 3.1-3.8 hold, the second-stage estimates $\widehat{\theta}_n(\tau)$ are asymptotically normal

$$\sqrt{n}\left(\widehat{\theta}_n(\tau) - \theta_0(\tau)\right) \rightsquigarrow \mathcal{N}\left(0, \boldsymbol{\Gamma}_{\theta,0}^{-1}\mathbb{E}\left[\mathbf{M}_t \Xi^{\tau} \mathbf{M}_t\right]\boldsymbol{\Gamma}_{\theta,0}^{-1}\right),$$

where

$$\begin{aligned}
\mathbf{M}_t = \mathbf{D}_2 \Bigg[& z_t(\alpha_0) G_t(\zeta_0), z_t(\alpha_0)(1 - G_t(\zeta_0)), \theta_0^{\Delta T} z_t(\alpha_0) \frac{\partial G_t(\zeta_0)}{\partial \zeta}\\
&, z_t^m G_t(\zeta_0), z_t^m (1 - G_t(\zeta_0)), \alpha_0^{\Delta} z_t^m \frac{\partial G_t(\zeta_0)}{\partial \zeta}\Bigg],
\end{aligned} \tag{3.12}$$

$\mathbf{D}_2 = \left[\mathbf{I}_{2(p+q+1)}, \frac{1}{\sum_{k=1}^{K} s_k q_k^2}\Gamma_{\alpha,0}\mathbf{D}_n^{-1}\right]$, and with a typical element of Ξ^{τ} defined as

$$\Xi_{i,j} = \begin{cases} \tau(1-\tau)/f_{\varepsilon}(F_{\varepsilon}^{-1}(\tau))^2 & \text{if } i \leqslant p+q+1 \text{ and } j \leqslant p+q+1 \\ q_i q_j (\tau_i \wedge \tau_j)(1 - \tau_i \vee \tau_j)/[f_{\varepsilon}(F_{\varepsilon}^{-1}(\tau_i)) f_{\varepsilon}(F_{\varepsilon}^{-1}(\tau_j))] & \text{if } i > p+q+1 \text{ and } j > p+q+1\,. \\ q_i(\tau_i \wedge \tau)(1 - \tau_i \vee \tau)/[f_{\varepsilon}(F_{\varepsilon}^{-1}(\tau_i)) f_{\varepsilon}(F_{\varepsilon}^{-1}(\tau))] & \text{otherwise } (i > j \text{ w.l.o.g}). \end{cases} \tag{3.13}$$

Proof. The first order condition $g_n(\theta(\tau), \widehat{\alpha}_n) = 0$ is solved by $\widehat{\theta}_n(\tau)$ such that

$$0 = g_n\left(\widehat{\theta}_n(\tau), \widehat{\alpha}_n\right) = g_n(\theta_0, \alpha_0) + \Gamma_{\theta,0}(\widehat{\theta}_n(\tau) - \theta_0(\tau)) + \Gamma_{\alpha,0}(\widehat{\alpha}_n - \alpha_0) + o_p\left(n^{-\frac{1}{2}}\right),$$

using the linearization from Corollary 3.2. Since $\Gamma_{\theta,0}$ has full rank, by premultiplying $\sqrt{n}$, an asymptotic representation of the second stage estimator is obtained

$$\sqrt{n}\left(\widehat{\theta}_n(\tau) - \theta_0(\tau)\right) = -\boldsymbol{\Gamma}_{\theta,0}^{-1}\left[\sqrt{n}g_n(\theta_0(\tau), \alpha_0)) + \boldsymbol{\Gamma}_{\alpha,0}\sqrt{n}(\widehat{\alpha}_n - \alpha_0)\right] + o_p(1).$$

where we use Assumption 3.5 and the α-mixing central limit theorem [Ibragimov & Linnik, 1971, Theorem 18.5.3], for which we require $(2+\delta)$ moments of our data and mixing coefficient satisfying $\beta_s \to 0$ and $\sum_{s=1}^{\infty} \beta_s^{\delta/(2+\delta)} < +\infty$.

We can then stack the two summands and write the last equation as

$$\sqrt{n}\left(\widehat{\theta}_n(\tau) - \theta_0(\tau)\right) = \Gamma_{\theta,0}^{-1}\left[\mathbf{I}_{2(p+q+1)}, \Gamma_{\alpha,0}\right]\sqrt{n}\begin{bmatrix} g_n(\theta_0(\tau), \alpha_0) \\ \widehat{\alpha_n^{\mathrm{I}}} - \alpha_0^{\mathrm{I}} \\ \widehat{\alpha_n^{\mathrm{II}}} - \alpha_0^{\mathrm{II}} \\ \widehat{\zeta_n} - \zeta_0 \end{bmatrix} + o_P(1)$$

$$\approx \frac{1}{\sqrt{n}}\Gamma_{\theta,0}^{-1}\left[\mathbf{I}_{2(p+q+1)}, \frac{1}{\sum_{k=1}^{K} s_k q_k^2}\Gamma_{\alpha,0}\mathbf{D}_n^{-1}\right]\sum_{t=m+1}^{T}\begin{bmatrix} \begin{bmatrix} z_t(\alpha_0)G_t(\zeta_0) \\ z_t(\alpha_0)(1 - G_t(\zeta_0)) \\ \theta_0^{\Delta T} z_t(\alpha_0)\frac{\partial G_t(\zeta_0)}{\partial \zeta} \end{bmatrix}\left(\mathbb{1}\left\{u_t \leqslant F_{u_t|\mathcal{F}_{t-1}}^{-1}(\tau)\right\} - \tau\right) \\ \begin{bmatrix} G_t(\zeta_0)\, z_t^m \\ (1 - G_t(\zeta_0)) z_t^m \\ \alpha_0^{\Delta} z_t^m \frac{\partial G_t(\zeta_0)}{\partial \zeta} \end{bmatrix}\sum_{k=1}^{K} q_k\left(\mathbb{1}\left\{u_t \leqslant F_{u_t|\mathcal{F}_{t-1}}^{-1}(\tau_k)\right\} - \tau_k\right) \end{bmatrix}$$

$$\rightsquigarrow \mathcal{N}\left(0, \boldsymbol{\Gamma}_{\theta,0}^{-1}\mathbb{E}\left[\mathbf{M}_t \Xi^{\tau} \mathbf{M}_t^{T}\right]\boldsymbol{\Gamma}_{\theta,0}^{-1}\right)$$

where we use independence of the innovations ε_t and z_t^m and $z_t(\alpha_0)$, respectively. The matrices Ξ^{τ}, $\mathbf{M}_t$, and $\boldsymbol{\Gamma}_{\theta,0}$ are defined in equations (3.13), (3.12) and (3.28).

□

Bibliography

ADAMS, A., CHERCHYE, L., DE ROCK, B. & VERRIEST, E. (2014). Consume now or later? Time inconsistency, collective choice, and revealed preference. *American Economic Review* 104 4027–4070.

AFRIAT, S. (1967). The construction of utility functions from expenditure data. *International Economic Review* 8 67–77.

ALTONJI, J. & MATZKIN, R. (2005). Cross section and panel data estimators for nonseparable models with endogenous regressors. *Econometrica* 73 1053–1102.

ANDERSEN, T. & BOLLERSLEV, T. (2006). Volatility and correlation forecasting. *Journal of Economic Forecasting* 1 777–878.

ANDERSON, H. M., NAM, K. & VAHID, F. (1999). Asymmetric nonlinear smooth transition GARCH models. *Dynamic Modeling and Econometrics in Economics and Finance* 1 191–207.

ANDREWS, D. W. (1994). Empirical Process Methods in Econometrics. *Handbook of Econometrics* IV 2248–2294.

ANDREWS, D. W. K. (2000). Inconsistency of the Bootstrap when a Parameter is on the Boundary of the Parameter Space. *Econometrica* 68 399–405.

BARMBY, T. & SMITH, N. (2001). Household labour supply in Britain and Denmark: some interpretations using a model of Pareto Optimal behaviour. *Applied Economics* 33 1109–1116.

BEATTY, T. K. M. & CRAWFORD, I. A. (2011). How demanding is the revealed preference approach to demand? *American Economic Review* 101 2782–2795.

BECKER, G. S. (1981). Altruism in the Family and Selfishness in the Market Place. *Economica* 48 1–15.

BECKERT, W. & BLUNDELL, R. (2008). Heterogeneity and the non-parametric analysis of consumer choice: Conditions for invertibility. *The Review of Economic Studies* 75 1069–1080.

BENKARD, C. L. & BERRY, S. (2006). On the nonparametric identification of nonlinear simultaneous equations models: Comment. *Econometrica* 74 1429–1440.

BLUNDELL, R., CHIAPPORI, P., MAGNAC, T. & MEGHIR, C. (2007). Collective labour supply: Heterogeneity and non-participation. *The Review of Economic Studies* 74 417–445.

BLUNDELL, R., HOROWITZ, J. & PAREY, M. (2013). Nonparametric estimation of a heterogeneous demand function under the Slutsky inequality restriction. *CEMMAP Working Paper* CWP54.

BLUNDELL, R., KRISTENSEN, D. & MATZKIN, R. (2014). Bounding quantile demand functions using revealed preference inequalities. *Journal of Econometrics* 179 112–127.

BLUNDELL, R. & POWELL, J. (2003). Endogeneity in Nonparametric and Semiparametric Regression Models. *Econometric Society Monographs* 8 312–357.

BLUNDELL, R. & STOKER, T. (2007). Models of aggregate economic relationships that account for heterogeneity. *Handbook of Econometrics* Volume 6a 4610–4666.

BOLLERSLEV, T. (1986). Generalized Autoregressive Conditional Heteroskedasticity. *Journal of Econometrics* 31 307–327.

BONKE, J. & BROWNING, M. (2015). Pooling of income and sharing of consumption within households. *Review of Economics of the Household* 13 73–93.

BROWN, B. (1983). The identification problem in systems nonlinear in the variables. *Econometrica* 51 175–196.

BROWN, B. & WALKER, M. (1989). The random utility hypothesis and inference in demand systems. *Econometrica* 57 815–829.

BROWNING, M. & CHIAPPORI, P. (1998). Efficient Intra-Household Allocations: A General Characterization and Empirical Tests. *Econometrica* 66 1241–1278.

BROWNING, M., CHIAPPORI, P. & LEWBEL, A. (2013). Estimating consumption economies of scale, adult equivalence scales, and household bargaining power. *Review of Economic Studies* 80 1267–1303.

BROWNING, M. & GORTZ, M. (2012). Spending Time and Money within the Household. *The Scandinavian Journal of Economics* 114 681–704.

CAI, Z. & XU, X. (2009). Nonparametric Quantile Estimations for Dynamic Smooth Coefficient Models. *Journal of the American Statistical Association* 104 371–383.

CARRASCO, M. & CHEN, X. (2002). Mixing and moment properties of various GARCH and stochastic volatility models. *Econometric Theory* 18 17–39.

CHAN, F. & MCALEER, M. (2003). Estimating smooth transition autoregressive models with GARCH errors in the presence of extreme observations and outliers. *Applied Financial Economics* 13 581–592.

CHAUDHURI, P. (1991). Nonparametric estimates of regression quantiles and their local Bahadur representation. *The Annals of Statistics* 19 760–777.

CHEN, X. (2008). Large Sample Sieve Estimation of Semi-Nonparametric Models. In J. J. Heckman & E. E. Leamer, eds., *Handbook of Econometrics*. Elsevier, volume 6b ed., 5549–5633.

CHEN, X., LINTON, O. & KEILEGOM, I. V. (2003). Estimation of semiparametric models when the criterion function is not smooth. *Econometrica* 71 1591–1608.

CHEN, X. & SHEN, X. (1998). Sieve Extremum Estimates for Weakly Dependent Data. *Econometrica* 66 289–314.

CHERCHYE, L., DE ROCK, B. & VERMEULEN, F. (2009). Opening the black box of intrahousehold decision-making. *Journal of Political Economy* 117 1074–1104.

CHERCHYE, L., DE ROCK, B. & VERMEULEN, F. (2011). The Revealed Preference Approach to Collective Consumption Behaviour: Testing and Sharing Rule Recovery. *The Review of Economic Studies* 78 176–198.

CHERCHYE, L., DE ROCK, B. & VERMEULEN, F. (2012). Married with children: A collective labor supply model with detailed time use and intrahousehold expenditure information. *American Economic Review* 102 3377–3405.

CHERCHYE, L., ROCK, B. D., LEWBEL, A. & VERMEULEN, F. (2015). Sharing rule identification for general collective consumption models. *Econometrica* 83 2001–2041.

CHERCHYE, L., ROCK, B. D. & VERMEULEN, F. (2007). The collective model of household consumption: a nonparametric characterization. *Econometrica* 75 553–574.

CHESHER, A. (2003). Identification in nonseparable models. *Econometrica* 71 1405–1441.

CHESHER, A. & ROSEN, A. M. (2014). Generalized Instrumental Variable Models. *CEMMAP Working Paper* 43 1–53.

CHIAPPORI, P. (1988). Rational household labor supply. *Econometrica* 56 63–90.

CHIAPPORI, P. (1992). Collective labor supply and welfare. *Journal of Political Economy* 100 437–467.

CHIAPPORI, P., DONNI, O. & KOMUNJER, I. (2012). Learning from a Piece of Pie. *The Review of Economic Studies* 79 162–195.

CHIAPPORI, P. & EKELAND, I. (2006). The micro economics of group behavior: General characterization. *Journal of Economic Theory* 130 1–26.

CHIAPPORI, P. & EKELAND, I. (2009). The Microeconomics of Efficient Group Behavior: Identification. *Econometrica* 77 763–799.

CHIAPPORI, P. & KIM, J. (2013). Identifying Heterogeneous Sharing Rules. *Columbia University* .

DEATON, A. & MUELLBAUER, J. (1980). An almost ideal demand system. *American Economic Review* 70 312–326.

DETTE, H., HODERLEIN, S. & NEUMEYER, N. (2016). Testing multivariate economic restrictions using quantiles: the example of Slutsky negative semidefiniteness. *Journal of Econometrics* 191 129–144.

DOORNIK, J. (2009). *An Object-Oriented Matrix Programming Language Ox 6.* London: Timberlake Consultants Ltd.

DUNBAR, G., LEWBEL, A. & PENDAKUR, K. (2013a). Children's resources in collective households: identification, estimation, and an application to child poverty in Malawi. *American Economic Review* 103 438–471.

DUNBAR, G., LEWBEL, A. & PENDAKUR, K. (2013b). Identification of random resource shares in collective households with an application to Microcredit in Malawi. *Boston College* .

ENGLE, R. (1982). Autoregressive conditional heteroscedasticity with estimates of the variance of United Kingdom inflation. *Econometrica* 50 987–1007.

ENGLE, R. F. & MANGANELLI, S. (2004). CAViaR: Conditional Autoregressive Value at Risk by Regression Quantiles. *Journal of Business and Economic Statistics* 22 367–381.

EVDOKIMOV, K. (2010). Identification and Estimation of a Nonparametric Panel Data Model with Unobserved Heterogeneity. *Princeton University: Discussion Paper* 1–62.

FISHER, F. (1966). *The identification problem in econometrics*. New York: McGraw-Hill.

FORTIN, B. & LACROIX, G. (1997). A test of the unitary and collective models of household labour supply*. *The Economic Journal* 107 933–955.

FRANC, V., HLAVAC, V. & NAVARA, M. (2005). Sequential coordinate-wise algorithm for the non-negative least squares problem. *Computer Analysis of Images and Patterns* 3691 407–414.

FRÉCHET, M. (1951). Sur les tableaux de corrélation dont les marges sont données. *Annals de l'Université de Lyon* 9 53–77.

GERLACH, R. H., CHEN, C. W. S. & CHAN, N. Y. C. (2011). Bayesian Time-Varying Quantile Forecasting for Value-at-Risk in Financial Markets. *Journal of Business & Economic Statistics* 29 481–492.

GLOSTEN, L. R., JAGANNATHAN, R. & RUNKLE, D. E. (1993). On the relation between the expected value and the volatility of the nominal excess return on stocks. *Journal of Finance* 48 1779–1801.

GONZALES-RIVERA, G. (1998). Smooth-Transition GARCH Models. *Studies in Nonlinear Dynamics and Econometrics* 3 61–78.

GORMAN, W. (1953). Community preference fields. *Econometrica* 21 63–80.

GORMAN, W. (1961). On a class of preference fields. *Metroeconomica* 13 53–56.

HAAG, B. R., HODERLEIN, S. & PENDAKUR, K. (2009). Testing and imposing Slutsky symmetry in nonparametric demand systems. *Journal of Econometrics* 153 33–50.

HÄRDLE, W., HILDENBRAND, W. & JERISON, M. (1991). Empirical evidence on the law of demand. *Econometrica* 59 1525–1549.

HÄRDLE, W. & STOKER, T. M. (1989). Investigating Smooth Multiple Regression by the Method of Average Derivatives. *Journal of the American Statistical Association* 84 986–995.

HÄRDLE, W. K. & SONG, S. (2009). Confidence Bands in Quantile Regression. *Econometric Theory* 26 1180–1200.

HODERLEIN, S. (2011). How many consumers are rational? *Journal of Econometrics* 164 294–309.

HODERLEIN, S. & MAMMEN, E. (2007). Identification of marginal effects in nonseparable models without monotonicity. *Econometrica* 75 1513–1518.

HODERLEIN, S. & MAMMEN, E. (2009). Identification and estimation of local average derivatives in non-separable models without monotonicity. *The Econometrics Journal* 12 1–25.

HODERLEIN, S. & STOYE, J. (2014). Revealed Preferences in a Heterogeneous Population. *Review of Economics and Statistics* 96 197–213.

HUANG, D., YU, B., FABOZZI, F. J. & FUKUSHIMA, M. (2009). CAViaR-based forecast for oil price risk. *Energy Economics* 31 511–518.

HUBER, P. J. (1964). Robust Estimation of a Location Parameter. *The Annals of Mathematical Statistics* 35 73–101.

HUBNER, S. (2015). Nonparametric Identification and Estimation of the Sharing Rule in Collective Models with Unobserved Heterogeneity. *Discussion paper: Tilburg University* .

HURWICZ, L. (1950). *Generalization of the Concept of Identification*, vol. 10. Cowles Commission.

HURWICZ, L. & UZAWA, H. (1971). On the integrability of demand functions. In *Preferences, utility and demand*. New York, Chapter 6.

IBRAGIMOV, I. & LINNIK, J. (1971). *Independent and stationary sequences of random variables*.

IMBENS, G. & NEWEY, W. (2009). Identification and estimation of triangular simultaneous equations models without additivity. *Econometrica* 77 1481–1512.

JOHANSSON, B., ELFVING, T., KOZLOV, V., CENSOR, Y., FORSSEN, P. & GRANLUND, G. (2006). The application of an oblique-projected Landweber method to a model of supervised learning. *Mathematical and Computer Modelling* 43 892–909.

KITAMURA, Y. & STOYE, J. (2013). Nonparametric Analysis of Random Utility Models: Testing. *Cowles Foundation Discussion Paper* 1902 1–48.

KLAASSEN, F. (2002). Improving GARCH volatility forecasts with regime-switching GARCH. *Empirical Economics* 27 363–394.

KOENKER, R. (2000). *Quantile Regression*. February 1997.

KOENKER, R. (2005). *Quantile regression*. Cambridge: Econometric Society Monographs, 1st ed.

KOENKER, R. & BASSETT, G. (1978). Regression quantiles. *Econometrica* 46 33–50.

KOENKER, R. & ZHAO, Q. (1996). Conditional quantile estimation and inference for ARCH models. *Econometric Theory* 12 793–813.

KOOPMANS, T. & REIERSOL, O. (1950). The identification of structural characteristics. *The Annals of Mathematical Statistics* 21 165–181.

KOSOROK, M. R. (2008). *Introduction to Empirical Processes and Semiparametric Inference*. August. New York: Springer.

KUPIEC, P. (1995). Techniques for verifying the accuracy of risk measurement models. *The Journal of Derivatives* 3 73–84.

LANNE, M. (2005). Non-linear GARCH models for highly persistent volatility. *The Econometrics Journal* 8 251–276.

LAWSON, C. & HANSON, R. (1995). Solving Least Squares Problems. In *SIAM*. Classics i ed.

LEE, S. & NOH, J. (2013). Quantile Regression Estimator for GARCH Models. *Scandinavian Journal of Statistics* 40 2–20.

LEWBEL, A. (2001). Demand Systems with and without Errors. *American Economic Review* 31 611–618.

LEWBEL, A. (2016). The Identification Zoo - Meanings of Identification in Econometrics. *Boston College Working Paper* 1–80.

LI, C. W. & LI, W. K. (1996). On a double-threshold autoregressive heteroscedastic time series model. *Journal of Applied Econometrics* 11 253–274.

LUNDBERG, S., POLLAK, R. & WALES, T. (1997). Do husbands and wifes pool their resources? *Journal of Human Resources* 32 463–480.

LUUKKONEN, R., SAIKKONEN, P. & TERÄSVIRTA, T. (1988). Testing linearity against smooth transition autoregressive models. *Biometrika* 75 491–499.

MANSKI, C. F. (1995). *Identification problems in the social sciences*. Cambridge: Harvard University Press.

MANSKI, C. F. (2003). *Partial Identification of Probability Distributions*. New York: Springer.

MANSKI, C. F. (2007). Partial Identification of Counterfactual Choice Probabilities. *International Economic Review* 48 1393–1410.

MATZKIN, B. R. L. (2008). Identification in Nonparametric Simultaneous Equations Models. *Econometrica* 76 945–978.

MATZKIN, R. (2003). Nonparametric estimation of nonadditive random functions. *Econometrica* 71 1339–1375.

MATZKIN, R. & PEREZ-ESTRADA, J. (2011). Heterogeneous Collective Demand: Exploring Differential Generosity within Elderly Couples. *Discussion paper: UCLA* .

MCFADDEN, D. (2005). Revealed stochastic preference: a synthesis. *Economic Theory* 26 245–264.

MCFADDEN, D. & RICHTER, M. (1991). Stochastic rationality and revealed stochastic preference. In J. Chipman, D. McFadden & M. Richter., eds., *Preferences, Uncertainty, and Optimality*. Boulder: Westview Press, 161–186.

MCQUEEN, G. & VORKINK, K. (2004). Whence GARCH? A Preference-Based Explanation for Conditional Volatility. *The Review of Financial Studies* 17 915–949.

MEITZ, M. & SAIKKONEN, P. (2008). Ergodicity, mixing, and existence of moments of a class of Markov models with applications to GARCH and ACD models. *Econometric Theory* 24 1291–1320.

MENON, M., PENDAKUR, K. & PERALI, F. (2012). On the expenditure-dependence of children's resource shares. *Economics Letters* 117 739–742.

NAM, K., PYUN, C. S. & AVARD, S. L. (2001). Asymmetric reverting behavior of short-horizon stock returns: An evidence of stock market overreaction. *Journal of Banking & Finance* 25 807–824.

NEWEY, W. K. & MCFADDEN, D. (1994). Large Sample Estimation and Hypothesis Testing. *Handbook of Econometrics* 4 2112–2245.

POPPER, K. (1934). *Logik der Forschung. Zur Erkenntnistheorie der modernen Naturwissenschaft*, vol. 1. Vienna: Springer.

POWELL, J. L. (1994). Chapter 41 Estimation of semiparametric models. *Handbook of Econometrics* 4 2443–2521.

ROEHRIG, C. (1988). Conditions for identification in nonparametric and parametric models. *Econometrica* 56 433–447.

ROTHENBERG, T. (1971). Identification in Parametric Models. *Econometrica* 39 577–591.

SABBAH, C. (2014). Uniform Confidence Bands for Local Polynomial Quantile Estimators. *ESAIM: Probability and Statistics* 18 265–276.

SCHWERT, G. (1990). Stock Volatility and the Crash of'87. *Review of Financial Studies* 3 77–102.

TAMER, E. (2010). Partial Identification in Econometrics. *Annual Review of Economics* 2 167–195.

TAYLOR, S. (1986). *Modelling Financial Time Series*. Chichester: Wiley.

TERASVIRTA, T. (1992). Characterizing nonlinearities in business cycles using smooth transition autoregressive models. *Journal of Applied Econometrics* 7 119–136.

THOMAS, D. (1990). Intra-household resource allocation: An inferential approach. *Journal of Human Resources* 25 635–664.

TONG, H. & LIM, K. (1980). Threshold autoregression, limit cycles and cyclical data. *Journal of the Royal Statistical Society. Series B (Methodological)* 42 245–292.

VAN DER VAART, A. & WELLNER, J. (1996). *Weak Convergence and Empirical Processes: With Applications to Statistics*. New York: Springer.

VARIAN, H. (1982). The nonparametric approach to demand analysis. *Econometrica* 50 945–73.

VERMEULEN, F., BARGAIN, O., BEBLO, M., BENINGER, D., BLUNDELL, R., CARRASCO, R., CHIURI, M.-C., LAISNEY, F., LECHENE, V., MOREAU, N., MYCK, M. & RUIZ-CASTILLO, J. (2006). Collective Models of Labor Supply with Nonconvex Budget Sets and Nonparticipation: A Calibration Approach. *Review of Economics of the Household* 4 113–127.

WELSH, A. (1989). On M-processes and M-estimation. *The Annals of Statistics* 17 337–361.

WHITE, H., TAE-HWAN, K. & MANGANELLI, S. (2008). Modeling autoregressive conditional skewness and kurtosis with multi-quantile CAViaR. *ECB Working Paper* 957.

XIAO, Z., GUO, H. & LAM, M. S. (2015). *Handbook of Financial Econometrics and Statistics*. New York, NY: Springer New York.

XIAO, Z. & KOENKER, R. (2009). Conditional quantile estimation for generalized autoregressive conditional heteroscedasticity models. *Journal of the American Statistical Association* 104 1696–1712.

YU, K. & JONES, M. (1998). Local linear quantile regression. *Journal of the American Statistical Association* 93 228–237.

ZHAO, Z. & XIAO, Z. (2014). Efficient Regressions via Optimally Combining Quantile Information. *Econometric theory* 30 1272–1314.

ZHENG, S. (2011). Gradient descent algorithms for quantile regression with smooth approximation. *International Journal of Machine Learning and Cybernetics* 2 191–207.

ZOU, H. & YUAN, M. (2008). Composite quantile regression and the oracle model selection theory. *The Annals of Statistics* 36 1108–1126.